

A Logic of Objective and Subjective Oughts

Aldo Iván Ramírez Abarca¹ and Jan Broersen²

¹ Utrecht University, Utrecht 3512 JK, The Netherlands a.i.ramirezabarca@uu.nl

² Utrecht University, Utrecht 3512 JK, The Netherlands J.M.broersen@uu.nl

Abstract. The relation between agentive action, knowledge, and obligation is central to the understanding of responsibility –a main topic in Artificial Intelligence. Based on the view that an appropriate formalization of said relation would contribute to the development of ethical AI, we point out the main characteristics of a logic for objective and subjective oughts that was recently introduced in the literature. This logic extends the traditional stit paradigm with deontic and epistemic operators, and provides a semantics that deals with Horty’s puzzles for knowledge and obligation. We provide an axiomatization for this logic, and address its soundness and completeness with respect to a class of relevant models.

1 Introduction

AI developers face a big challenge in creating systems that are expected to make ethically charged decisions. The field of machine ethics has seen a quick growth in recent years, and questions regarding *responsibility* of autonomous agents are very important. In our opinion, these questions can be categorized in two main trends: 1) *conceptual* questions that deal with the ontology and essential components of the notion of responsibility, and 2) *technical* questions concerning the implementation of such notion in AI. This work attempts to make a contribution in the *technical* category. We take part in a very specific line of research, where proof systems of deontic logic are intended to help in the testing of ethical behavior of AI through theorem proving ([1], [9], [3]). That being said, answers to the technical kind of questions typically presuppose a particular choice of philosophical standpoint against questions of the first kind, the conceptual one. Our philosophical standpoint comes from John Horty’s framework of act utilitarian stit logic ([8]), extended with epistemic relations. As such, we make explicit the goal of the present paper: to provide well-behaved formalizations of 3 essential components of responsibility of intelligent systems: *agentive action*, *knowledge*, and *obligation*.

According to [1], having an *expressive deontic logic with practical relevance* and an efficient algorithm for proving theorem-hood is highly applicable in the construction of logic-based ethical robots.³ To support this statement, Arkoudas et al. present in [1] a natural deduction calculus for a logic of ought-to-do that was developed by Horty in [8] and axiomatised by Murakami in [9]. With an interactive theorem proving system named ATHENA, they illustrate the fact that we can mechanize deontic logic to do machine ethics.

³ See [10] for an overview of the advantages and disadvantages of doing machine ethics via theorem proving.

The starting point of our work comes from a recent interest in enhancing both the expressivity and practical relevance of Horty’s stit theory of ought-to-do in order to deal with situations in which agents’ *knowledge* becomes significant. Inspired by 3 puzzles for knowledge-dependent obligations that pose a problem for merely extending his initial logic with epistemic operators, Horty presents in [6] a novel semantics for *epistemic oughts* based on *action types*. Although the approach is substantial, it comes with 3 disadvantages: (1) it diverges from his work in [8], (2) there are semantic constraints that limit the expressivity of the models, and (3) the use of action types precludes an efficient axiomatization.⁴

With similar motivations as Horty, the extended abstract [5] proposes an alternative logic –where the main idea is to distinguish objective from subjective obligations– that allegedly mitigates the disadvantages mentioned above. The authors claim that this logic is simpler, more naturally connected to Horty’s work in [8], and axiomatizable. Being very brief, [5] only deals with the conceptual benefits of the proposal, and the proof of soundness and completeness of the logic’s proof system is only mentioned. Here, we recover the definition of this logic, address its benefits, and show that its proof system is indeed sound and complete with respect to a class of relevant models. We do this hoping that the results will give some background to new developments in the mechanization of deontic logic for ethical AI, in the aforementioned tradition of [1] and [3].

The paper is structured as follows. After a short presentation of stit and its applications in the modeling of action, knowledge, and obligation, we go through Horty’s puzzles. We mention his solution to them and justify the claim that his approach with action types comes with disadvantages. Afterwards, we present the logic developed by [5], show how it solves Horty’s puzzles, and deal with its axiomatizability.⁵

2 Action, Knowledge, and Obligation in Stit

We want to consider oughts from the perspective that an agent should be excused for having failed at an obligation if it lacks the necessary knowledge to perform a required task (*doctors ought to stop the bleeding of the patient, but if they do not know how to, they should be excused*). A typical framework for expressing statements that involve knowledge for required tasks as a component of responsibility is stit logic ([2]). Stit logic was created to formalize the concept of *action*, so it naturally lent itself to the study of *obligation* ([8]) and of *knowingly doing* ([4]), all important elements in the notion of responsibility. For a comprehensive review of the interaction of these 3 concepts in the literature, we refer the reader to [11].

The 3 puzzles of Horty that we mentioned in the introduction are actually very good illustrations of how stit deals with action, knowledge, and obligation, but in order to tackle them we need to recover basic definitions of the paradigm known as *act utilitarian stit logic*. In this paradigm, *obligation* stems from a dominance ordering over the set of available actions. The idea is that the effects of the best actions in the ordering –the

⁴ We will justify all these claims in the second section of this paper, after introducing the 3 puzzles and addressing Horty’s solution.

⁵ The proof of completeness is dense and technical, so the full proofs of each statement are provided in the appendix section of the present work.

so-called *optimal* actions—are the obligations of a given agent at a given moment. We proceed to introduce the basic aspects of an extension of act utilitarian stit logic with epistemic modalities, and leave its examples to the section where we present Horty’s puzzles.

Definition 1 (Syntax). *Given a finite set Ags of agent names, a countable set of propositions P such that $p \in P$ and $\alpha \in Ags$, the grammar for the formal language \mathcal{L}_{KO} is given by:*

$$\varphi := p \mid \neg\varphi \mid \varphi \wedge \psi \mid \Box\varphi \mid [\alpha]\varphi \mid K_\alpha\varphi \mid \odot[\alpha]\varphi$$

$\Box\varphi$ is meant to express the ‘historical necessity’ of φ ($\Diamond\varphi$ abbreviates $\neg\Box\neg\varphi$). $[\alpha]\varphi$ stands for ‘agent α sees to it that φ ’. K_α is the epistemic operator for α . Finally, $\odot[\alpha]\varphi$ is meant to represent that α ought to see to it that φ .

As for the semantics, the structures in which we evaluate formulas of the language \mathcal{L}_{KO} are based on what we call *epistemic act utilitarian branching time frames*.

Definition 2 (Branching time (BT) frames). *A **finite-choice** epistemic act utilitarian BT-frame is a tuple $\langle T, \sqsubset, \mathbf{Choice}, \{\sim_\alpha\}_{\alpha \in Ags}, \mathbf{Value} \rangle$ such that:*

- T is a non-empty set of moments and \sqsubset is a strict partial ordering on T satisfying ‘no backward branching’. Each maximal \sqsubset -chain is called a *history*, which represents a way in which time might evolve. H denotes the set of all histories, and for each $m \in T$, $H_m := \{h \in H; m \in h\}$. Tuples $\langle m, h \rangle$ are called *situations* iff $m \in T$, $h \in H$, and $m \in h$. **Choice** is a function that maps each agent α and moment m to a **finite** partition \mathbf{Choice}_α^m of H_m , where the cells of such a partition represent α ’s available actions at m . **Choice** satisfies two constraints:
 - (NC) or ‘no choice between undivided histories’.- For all $h, h' \in H_m$, if $m' \in h \cap h'$ for some $m' \sqsupset m$, then $h \in L$ iff $h' \in L$ for every $L \in \mathbf{Choice}_\alpha^m$.
 - (IA) or ‘independence of agency’.- A function s on Ags is called a *selection function* at m if it assigns to each α a member of \mathbf{Choice}_α^m . If we denote by \mathbf{Select}^m the set of all selection functions at m , then we have that for every $m \in T$ and $s \in \mathbf{Select}^m$, $\bigcap_{\alpha \in Ags} s(\alpha) \neq \emptyset$ (see [2] for a discussion of the property).
- For $\alpha \in Ags$, \sim_α is the epistemic indistinguishability equivalence relation for agent α .
- **Value** is a deontic function that assigns to each history $h \in H$ a real number, representing the utility of h .

As mentioned before, the idea is that obligations come from the optimal actions for a given agent. The optimality of such actions is relative to a dominance ordering of the actions, and this ordering is given by the value of the histories in those actions –itself provided by **Value**. In order to present the semantics for formulas involving the ought-to-do operator, we therefore need some previous definitions. For $\alpha \in Ags$ and $m_* \in T$, we define a dominance ordering \preceq on $\mathbf{Choice}_\alpha^{m_*}$ such that for $L, L' \in \mathbf{Choice}_\alpha^{m_*}$, $L \preceq L'$ iff $\mathbf{Value}(h) \leq \mathbf{Value}(h')$ for every $h \in L, h' \in L'$. We write $L \prec L'$ iff $L \preceq L'$ and $L' \not\preceq L$. The optimal set of actions, then, is taken as $\mathbf{Optimal}_\alpha^{m_*} := \{L \in \mathbf{Choice}_\alpha^{m_*}; \text{there is no } L' \in \mathbf{Choice}_\alpha^{m_*} \text{ such that } L \prec L'\}$. As is customary, the models and the semantics for the formulas are defined by adding a valuation function to the frames of Definition 2:

Definition 3. A BT-model \mathcal{M} consists of the tuple that results from adding a valuation function \mathcal{V} to a BT-frame, where $\mathcal{V} : P \rightarrow 2^{T \times H}$ assigns to each atomic proposition a set of moment-history pairs. Relative to a model \mathcal{M} , the semantics for the formulas of \mathcal{L}_{KO} is defined recursively by the following truth conditions, evaluated at a given situation $\langle m, h \rangle$:

$$\begin{aligned}
\langle m, h \rangle \models p & \quad \text{iff } \langle m, h \rangle \in \mathcal{V}(p) \\
\langle m, h \rangle \models \neg \varphi & \quad \text{iff } \langle m, h \rangle \not\models \varphi \\
\langle m, h \rangle \models \varphi \wedge \psi & \quad \text{iff } \langle m, h \rangle \models \varphi \text{ and } \langle m, h \rangle \models \psi \\
\langle m, h \rangle \models \Box \varphi & \quad \text{iff } \forall h' \in H_m, \langle m, h' \rangle \models \varphi \\
\langle m, h \rangle \models [\alpha] \varphi & \quad \text{iff } \forall h' \in \mathbf{Choice}_\alpha^m(h), \langle m, h' \rangle \models \varphi \\
\langle m, h \rangle \models K_\alpha \varphi & \quad \text{iff } \forall \langle m', h' \rangle \text{ s.t. } \langle m, h \rangle \sim_\alpha \langle m', h' \rangle, \langle m', h' \rangle \models \varphi \\
\langle m, h \rangle \models \odot[\alpha] \varphi & \quad \text{iff } \forall L \in \mathbf{Optimal}_\alpha^m, h' \in L \text{ implies that } \langle m, h' \rangle \models \varphi.
\end{aligned}$$

Satisfiability, validity on a frame, and general validity are defined as usual. We write $|\varphi|^m$ to refer to the set $\{h \in H_m; \mathcal{M}, \langle m, h \rangle \models \varphi\}$.

2.1 Horty's Puzzles

The 3 puzzles that we have mentioned since the beginning of the paper, and that pose a problem for formalizing epistemic oughts just with the epistemic extension of act utilitarian logic, can be summarized as follows.

Example 1. Agent β places a coin on top of a table –either heads up or tails up– but hides it from agent α . Agent α can bet that the coin is heads up, that it is tails up, or it can refrain from betting. If α bets and chooses correctly, it wins €10. If it chooses incorrectly, it does not win anything, and if it refrains from betting, it wins €5.

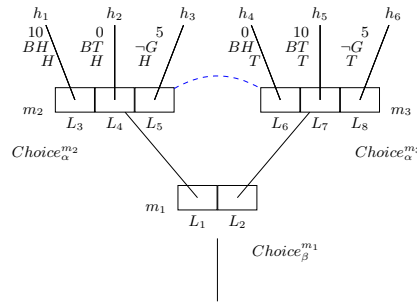


Fig. 1: Coin problem #1

The stit diagram that represents *Horty's interpretation* of the situation is included in Figure 1. In this diagram, we take H to denote the proposition ' β places the coin heads up', T to denote ' β places the coin tails up', BH to denote ' α bets heads', BT to denote ' α bets tails', and G to denote ' α gambles'. In moment m_1 , β places the coin on

top of the table, so that its available actions are labeled by L_1 (placing the coin heads up) and L_2 (placing the coin tails up). At moments m_2 and m_3 , it is α 's turn to act, and the available actions are clear from the picture. The blue dotted line represents the epistemic class of α : since β is hiding the coin, α cannot distinguish whether it is at moment m_2 or m_3 .⁶ For such an interpretation regarding the epistemic structure of the agent, a problem ensues due to the fact that for every $i \in \{2, 3\}$ and $h \in H_{m_i}$, we have that $\langle m_i, h \rangle \models K_\alpha \odot [\alpha]G$. This means that α knows that it ought to gamble, even if this is a 'risky' move that could result in a payoff of 0. In this sense, we could say that the agent's knowledge of what is optimal would lead it into taking a chance and gambling.

Example 2. With the same scheme as in the previous example, if α bets and chooses correctly, it wins €10. If it refrains from betting, it *also* wins €10. If it bets incorrectly, it does not win anything.

Intuitively, α ought to refrain from gambling in this scenario, for refraining implies that it would win by the same amount as when betting correctly but without engaging in an action that could possibly fail. In this case, the problem is that for every $i \in \{2, 3\}$ and $h \in H_{m_i}$, we have that $\langle m_i, h \rangle \not\models K_\alpha \odot [\alpha]\neg G$: α does not know that it ought to refrain from gambling. Figure 2 includes the stit diagram for this scenario.

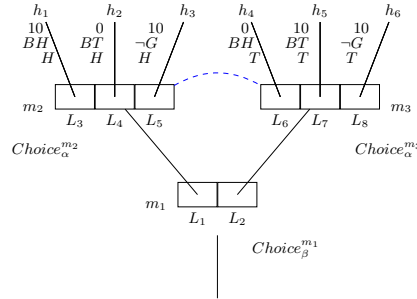


Fig. 2: Coin problem #2

Example 3. With the same scheme as in the previous examples, if α bets and chooses correctly, it wins €10. If it bets incorrectly or refrains from betting, it does not win anything.

The problem here is that for every $i \in \{2, 3\}$ and $h \in H_{m_i}$, $\langle m_i, h \rangle \models K_\alpha \odot [\alpha]W$, where W is the proposition that stands for ' α wins'. This means that α knows that it ought to win at any given situation, but such knowledge is not *action-guiding*, meaning that it will not provide α with a choice to make. Though the agent knowingly ought

⁶ Notice that Horty's formalization also yields that in none of the situations will α knowingly perform any of the available actions: it cannot epistemically distinguish between the situations in which it is 'betting heads', 'betting tails', or 'refraining'.

to win, it cannot knowingly do so –it simply does not have the means due to a lack of knowledge. Thus, Kant’s principle of ‘ought implies can’ is not satisfied ($\langle m_i, h \rangle \not\models K_\alpha \odot [\alpha]W \rightarrow \Diamond K_\alpha [\alpha]W$). Figure 3 includes the stit diagram for this scenario.

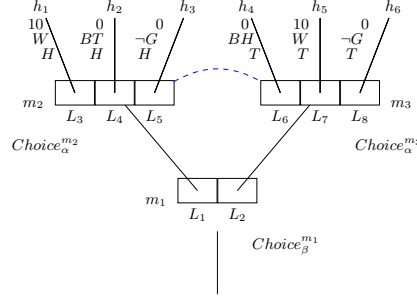


Fig. 3: Coin problem #3

Horty solves these 3 puzzles by introducing both syntactic and semantic addenda to epistemic act utilitarian stit logic. He extends the language with an operator $[... \text{kstit}]$ to encode the concept of *ex interim* knowledge (or knowingly doing). The semantics for formulas involving this operator uses *action types*, with the premise that actions of the same *type* may lead to different outcomes in different moments (see [7]). Unfortunately, the introduction of the new operator and of types comes with two unfavorable semantic constraints:

1. In order for $[... \text{kstit}]$ to be an **S5** operator, the epistemic relations must ensue not between moment-history pairs but between moments. The problem with this condition is that it limits the class of models to those in which knowledge is moment-dependent (agents will not be able to know that they perform a given action),⁷
2. Indistinguishable moments must offer the same available types. The problem with this constraint is that it cannot be characterized syntactically without producing an infinite axiomatization. This is due to the fact that performing a certain action type can only be expressed syntactically with propositional constants (again, see [7] for the details).⁸

As a solution to the stated puzzles, Horty’s approach is successful. However, [5] claims that we can also be successful without using action types. The benefits of the framework presented in [5], then, include technical characterizability of the constraints imposed on the structures –which is important for axiomatization–, semantic simplicity, and enhanced expressivity. In the following section, all these claims will be substantiated.

⁷ Horty’s models satisfy the following constraint: if $\langle m, h \rangle \sim_\alpha \langle m', h' \rangle$ then $\langle m, h' \rangle \sim_\alpha \langle m', h'' \rangle$ for every $h' \in H_m, h'' \in H_{m'}$.

⁸ It is an open problem to determine whether there is a finite axiomatization of Horty’s logic of epistemic action and obligation if the types were also included in the object language.

3 A Logic of Objective and Subjective oughts

[5] proposes to disambiguate two senses of ought-to-do in order to produce a system that solves Horty’s puzzles, avoids action types, and is axiomatizable. The two senses are an *objective* one, which coincides with Horty’s act utilitarian ought-to-do, and a *subjective* one, which arises from the epistemically best candidates in the set of available actions for a given agent. By ‘epistemically best’ we mean those actions that are undominated not only in the actual moment but whose all epistemic equivalents across different indistinguishable situations are also undominated.

Essentially, we are talking about an extension of the language in Definition 1 with a new operator $\odot_S[\alpha]$, meant to build up formulas that would express what α subjectively ought to do. As for the semantics of this new operator, it involves a dominance ordering as well, but one different to that which is used for objective ought-to-do’s. In order to define this subjective dominance ordering, [5] introduces a new semantic concept known as *epistemic clusters*, which are nothing more than a given action’s epistemic equivalents in situations that are indistinguishable to the actual one. Formally, we have that for $\alpha \in \text{Ags}$, $m_*, m \in T$, and $L \subseteq H_{m_*}$, L ’s *epistemic cluster* at m is the set

$$[L]_\alpha^m := \{h \in H_m; \exists h_* \in L \text{ s.t. } \langle m_*, h_* \rangle \sim_\alpha \langle m, h \rangle\}.$$

As a convention, we write $m \sim_\alpha m'$ if there exist $h \in H_m$, $h' \in H_{m'}$ such that $\langle m, h \rangle \sim_\alpha \langle m', h' \rangle$. The notion of epistemic clusters is used to define a subjective dominance ordering \preceq_s on $\text{Choice}_\alpha^{m_*}$ by the following rule: for $L, L' \in \text{Choice}_\alpha^{m_*}$, $L \preceq_s L'$ iff for every m such that $m_* \sim_\alpha m$, $\text{Value}(h) \leq \text{Value}(h')$ for every $h \in [L]_\alpha^m$, $h' \in [L']_\alpha^m$. Just as in the case of objective ought-to-do’s, this ordering allows us to define a subjectively optimal set of actions $\mathbf{S} - \text{optimal}_\alpha^{m_*} := \{L \in \text{Choice}_\alpha^{m_*}; \text{there is no } L' \in \text{Choice}_\alpha^{m_*} \text{ s.t. } L \prec_s L'\}$, where we write $L \prec_s L'$ iff $L \preceq_s L'$ and $L' \not\preceq_s L$. The idea, then, is that something will be a subjective obligation of a given agent at a given moment if it is an effect of all the subjectively optimal actions of that agent at that moment.

As established in [5], the models in which to evaluate the formulas of the extended language need to satisfy extra constraints in order to capture an appropriate interaction of action, knowledge, and subjective obligation. By this we mean to say that in these models (a) agents should be able to knowingly do the same things across epistemically indistinguishable states, (b) the subjective ought-to-do must conform to Kant’s directive of ‘ought implies can’ in its epistemic version of ‘subjectively ought-to-do implies ability of knowingly doing’, and (c) if something is a subjective ought-to-do of a given agent, then the agent should know that that is the case. Therefore, we focus on models that fulfill the following requirements, which will grant the conditions mentioned above (as can be seen from the proof of soundness):

- (OAC) For every situation $\langle m_*, h_* \rangle$, if $\langle m_*, h_* \rangle \sim_\alpha \langle m, h \rangle$ for some $\langle m, h \rangle$, then $\langle m_*, h'_* \rangle \sim_\alpha \langle m, h \rangle$ for every $h'_* \in \text{Choice}_\alpha^{m_*}(h_*)$. We refer to this constraint as the ‘own action condition’, since it implies that agents do not know more than what they perform. Because of this constraint, the knowledge that we are formalizing here is of a very particular kind: to know something is just the same as to knowingly do it.

- (Unif – H) For every situation $\langle m_*, h_* \rangle$, if $\langle m_*, h_* \rangle \sim_\alpha \langle m, h \rangle$ for some $\langle m, h \rangle$, then for every $h'_* \in H_{m_*}$, there exists $h' \in H_m$ such that $\langle m_*, h'_* \rangle \sim_\alpha \langle m, h' \rangle$. Combined with (OAC), this constraint is meant to capture a notion of uniformity of strategies, where epistemically indistinguishable situations should offer similar actions for the agent to choose upon. We call this condition ‘uniformity of historical possibility’.

In finite-choice epistemic act utilitarian BT-models that satisfy these two constraints, then, the semantics for the formulas involving $\odot_S[\alpha]$ is defined as expected:

$$\langle m, h \rangle \models \odot_S[\alpha]\varphi \text{ iff } \forall L \in \mathbf{S} - \mathbf{optimal}_\alpha^m, \forall m' \text{ s.t. } m \sim_\alpha m', [L]_\alpha^{m'} \subseteq |\varphi|^{m'}.$$

3.1 Solution to Horty’s Puzzles

The semantics for subjective ought-to-do’s offers solutions to natural interpretations of Horty’s puzzles, in which the assumption that the coin is hidden from the betting agent is captured by taking \sim_α to be defined by the following information sets: $\{\langle m_2, h_1 \rangle, \langle m_3, h_4 \rangle\}$, in which α bets heads; $\{\langle m_2, h_2 \rangle, \langle m_3, h_5 \rangle\}$, in which α bets tails; and $\{\langle m_2, h_3 \rangle, \langle m_3, h_6 \rangle\}$, in which α refrains from betting.

For Example 1, the problem is solved because although $\langle m_i, h_i \rangle \models K_\alpha \odot [\alpha]G$, we consider this as the knowledge of an *objective* ought-to-do. *Subjectively* speaking, we do not obtain that α knows that it ought to gamble: $\langle m_i, h_i \rangle \not\models \odot_S[\alpha]G$ and thus $\langle m_i, h_i \rangle \not\models K_\alpha \odot_S [\alpha]G$. This can be seen by noticing that $\mathbf{S} - \mathbf{Optimal}_\alpha^{m_2} = \{L_3, L_4, L_5\}$ and $\mathbf{S} - \mathbf{Optimal}_\alpha^{m_3} = \{L_6, L_7, L_8\}$. Figure 4 provides a stit diagram for this scenario.

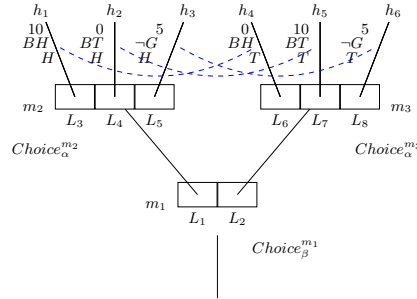


Fig. 4: Coin problem #1, revisited

In Example 2, the problem is solved because $\langle m_i, h_i \rangle \models \odot_S[\alpha]\neg G$ and $\langle m_i, h_i \rangle \models K_\alpha \odot_S [\alpha]\neg G$ (notice that $\mathbf{S} - \mathbf{Optimal}_\alpha^{m_2} = \{L_5\}$ and $\mathbf{S} - \mathbf{Optimal}_\alpha^{m_3} = \{L_8\}$).

For Example 3, the problem is solved because we obtain that $\langle m_i, h_i \rangle \not\models \odot_S[\alpha]W$, which in turn implies that $\langle m_i, h_i \rangle \not\models K_\alpha \odot_S [\alpha]W$ (notice that $\mathbf{S} - \mathbf{Optimal}_\alpha^{m_2} = \{L_3, L_4, L_5\}$ and $\mathbf{S} - \mathbf{Optimal}_\alpha^{m_3} = \{L_6, L_7, L_8\}$). Therefore, although α knows that it objectively ought to win, it is not the case that it subjectively ought to win.

When comparing these solutions to Horty's, it is important to point out that the formalization we are using is different from his, for it presupposes that the indistinguishability relation for agent α ensues between *situations*. Regardless of such difference, the solution is virtually the same: for Example 1, Horty gets that for every $h \in H_{m_2}$ $\langle m_2, h \rangle \not\models \odot[\alpha \text{ kstit}]G$. For Example 2, he gets that for every $h \in H_{m_2}$ $\langle m_2, h \rangle \models \odot[\alpha \text{ kstit}]\neg G$, and for Example 3, he gets that for every $h \in H_{m_2}$ $\langle m_2, h \rangle \not\models \odot[\alpha \text{ kstit}]W$. Therefore, we can see that the notion of $\odot_S[\alpha]$ works as an analog of $\odot[\dots \text{ kstit}]$.

4 Axiomatization and Some Logical Properties

In this section, we introduce a proof system for the logic presented, address its soundness and completeness results, and mention a few interesting properties of it.

Definition 4 (Proof system). *Let Λ be the proof system defined by the following axioms and rules of inference:*

(Axioms)

- All classical tautologies from propositional logic.
- The **S5** axiom schemata for \Box , $[\alpha]$, K_α ,⁹
- The following axiom schemata for the interactions of formulas with the given operators:

$$\odot[\alpha](p \rightarrow q) \rightarrow (\odot[\alpha]p \rightarrow \odot[\alpha]q) \quad (A1)$$

$$\Box p \rightarrow [\alpha]p \wedge \odot[\alpha]p \quad (A2)$$

$$\Box \odot[\alpha]p \vee \Box \neg \odot[\alpha]p \quad (A3)$$

$$\odot[\alpha]p \rightarrow \odot[\alpha](\odot[\alpha]p) \quad (A4)$$

$$\odot[\alpha]p \rightarrow \Diamond[\alpha]p \quad (Oic)$$

For $n \geq 1$ and pairwise different $\alpha_1, \dots, \alpha_n$,

$$\bigwedge_{1 \leq k \leq n} \Diamond[\alpha_k]p_k \rightarrow \Diamond \left(\bigwedge_{1 \leq k \leq n} [\alpha_k]p_k \right) \quad (IA)$$

$$\odot_S[\alpha](p \rightarrow q) \rightarrow (\odot_S[\alpha]p \rightarrow \odot_S[\alpha]q) \quad (A5)$$

$$\odot_S[\alpha]p \rightarrow \odot_S[\alpha](K_\alpha p) \quad (A6)$$

$$K_\alpha p \rightarrow [\alpha]p \quad (OAC)$$

$$\Diamond K_\alpha p \rightarrow K_\alpha \Diamond p \quad (Unif - H)$$

$$K_\alpha \Box p \rightarrow \odot_S[\alpha]p \quad (s.N)$$

$$\odot_S[\alpha]p \rightarrow \Diamond K_\alpha p \quad (s.Oic)$$

$$\odot_S[\alpha]p \rightarrow K_\alpha \Box \odot_S[\alpha]p \quad (Cl)$$

(Rules of inference)

- Modus Ponens, Substitution, and Necessitation for the modal operators.

⁹ The **S5** axiom schemata are standard from modal logic, and we include their names here just for coherence. They are (K), (T), (4), and (5).

Schema (IA) encodes ‘independence of agency’. (OAC) encodes the ‘own action condition’. $(Unif - H)$ encodes the ‘uniformity of historical possibility’ constraint. (Oic) and $(s.Oic)$ concern the objective, resp. subjective, versions of Kant’s directive of ‘ought implies can’. $(s.N)$ (standing for ‘subjective necessity’) captures that all that is historically necessary at epistemically indistinguishable situations must be a subjective obligation, and (Cl) (standing for ‘closure’) characterizes a property that says that if one subjectively ought to do something, then one knows that that is settled.

The axiom system \mathcal{A} turns out to be sound and complete with respect to the class of epistemic act utilitarian *bi-valued* BT-models. Such models are more general than the ones introduced in Definition 3. Instead of only one value function, there are two: one for the objective ought-to-do’s, and the other for the subjective ones.¹⁰ These models, then, are of the form $\langle T, \sqsubseteq, \mathbf{Choice}, \{\sim_\alpha\}_{\alpha \in Ags}, \mathbf{Value}_O, \mathbf{Value}_S \rangle$. As such, the models in Definition 3 are particular instances of bi-valued models, in which both value functions assign the same value to each history of the tree.

Furthermore, our soundness and completeness results presuppose a logic of ought-to-do that deals with the ‘sure-thing principle’, according to which the ranking of the available actions of a given agent should take into consideration what all the other agents choose concurrently. For a given agent, the action profiles of the other agents are seen as the possible states in which the agent will act (see [8], Chapter 4, subsection 4.1.2). We implement Horty’s approach for dealing with the ‘sure-thing principle’ in the logic that we axiomatize, and this means that the semantics that we use is in fact a generalization of the one introduced before. In order to address such a generalization and the soundness and completeness results, we need further definitions:

For $m \in T$ and $\beta \in Ags$, we define

$$\mathbf{State}_\beta^m = \{S \subseteq H_m; S = \bigcap_{\alpha \in Ags - \{\beta\}} s(\alpha), \text{ where } s \in \mathbf{Select}^m\}.$$

With this notion of states, we redefine the dominance orderings so that the actions are measured taking into consideration the states which those actions are facing. For $\alpha \in Ags$ and $m_* \in T$, we first define a general ordering \leq on $\mathcal{P}(H_{m_*})$ such that

¹⁰ This extension is extremely useful for the proof of completeness, but neither its conceptual reach nor its philosophical implications has been a subject of our investigation as of yet. There are more than a few reasons to entertain skepticism about this technical decision, since it is not grounded philosophically. To have two deontic functions instead of one lends itself to potentially having different values for histories according to whether their utility is objective or subjective. However, we must observe that the notions of objective and subjective utility are not what drives the disambiguation of ought-to-do into objective and subjective obligations. What drives it is the epistemic structure of a given agent. Therefore, the use of two value functions should not be taken as essential for the proposal of the logic presented here. Moreover, the soundness result works for models in which there is only one deontic function, and the solution to Horty’s puzzles does not rely on there being two. On a related note, we make explicit that the focus of this work was to develop well-behaved semantics for subjective ought-to-do’s in the stit tradition, so that if we presented a language with only the subjective-ought operator and not the objective one, as well as with the operators for knowledge, action, and historical necessity, then our proof of completeness would in fact be enough to show completeness of this fragment with respect to the class of single-valued epistemic act utilitarian BT-models.

for $X, Y \subseteq H_{m_*}$, $X \leq Y$ iff $\mathbf{Value}_O(h) \leq \mathbf{Value}_O(h')$ for every $h \in X, h' \in Y$. The objective dominance ordering \preceq is now defined such that for $L, L' \in \mathbf{Choice}_\alpha^{m_*}$, $L \preceq L'$ iff $\forall S \in \mathbf{State}_\alpha^{m_*}, L \cap S \leq L' \cap S$. As for the subjective ought-to-do's, we first define a general ordering \leq_s on $\mathcal{P}(H_{m_*})$ such that for $X, Y \subseteq H_{m_*}$, $X \leq_s Y$ iff $\mathbf{Value}_S(h) \leq_s \mathbf{Value}_S(h')$ for all $h \in X, h' \in Y$. The subjective dominance ordering \preceq_s is then defined such that for $L, L' \subseteq H_{m_*}$, $L \preceq_s L'$ iff $\forall m$ such that $m_* \sim_\alpha m$, $\forall S \in \mathbf{State}_\alpha^m, [L]_\alpha^m \cap S \leq_s [L']_\alpha^m \cap S$.

With these definitions, we adapt the semantics for the formulas involving ought-to-do operators, making the logic strong enough to deal with the ‘sure-thing principle’: we set $\langle m, h \rangle \models \odot[\alpha]\varphi$ iff for every $L \in \mathbf{Choice}_\alpha^m$ such that $\langle m, h_L \rangle \not\models \varphi$ for some $h_L \in L$, there exists $L' \in \mathbf{Choice}_\alpha^m$ such that $L \prec L'$ and if $L'' = L$ or $L' \preceq L''$, then $\langle m, h' \rangle \models \varphi$ for every $h' \in L''$. Similarly, we set $\langle m, h \rangle \models \odot_S[\alpha]\varphi$ iff for every $L \in \mathbf{Choice}_\alpha^m$ such that $\langle m', h_L \rangle \not\models \varphi$ for some $h_L \in [L]_\alpha^{m'}$ ($m \sim_\alpha m'$), there exists $L' \in \mathbf{Choice}_\alpha^m$ such that $L \prec_s L'$ and if $L'' = L'$ or $L' \preceq_s L''$, then for every m'' such that $m \sim_\alpha m''$, $\langle m'', h'' \rangle \models \varphi$ for every $h'' \in [L'']_\alpha^{m''}$.

4.1 Soundness

Proposition 1. *The system Λ is sound with respect to the class of epistemic act utilitarian bi-valued BT-models.*

Proof. Standard. See Appendix for details.

4.2 Completeness

For reasons of space, we can only summarize the proof here, leaving the full detailed exposition for the Appendix. The proof of completeness of Λ with respect to the class of epistemic act utilitarian bi-valued BT-models is a two-step process. First, we introduce relational structures called Kripke-estit models for evaluating the formulas of the language \mathcal{L}_{KO} under a semantics that mirrors the semantics of Definition 3, and prove completeness of Λ with respect to these structures via the technique of canonical models. Secondly, we provide a truth-preserving correspondence between Kripke-estit models and certain epistemic act utilitarian bi-valued BT-models, so that completeness with respect to Kripke-estit models grants it with respect to bi-valued BT-models. Essential to the success of the technique of canonical models is to prove the so-called *existence* and *truth* lemmas just as in modal logic. In the case of this particular proof system, these lemmas are quite involved, and the reader is encouraged to go over them in the Appendix, which offers enough room to address all their features carefully and with precision. In our opinion, the results of soundness and completeness are significant, all the more because we want to provide some theoretical foundations to the idea –explored by [1]– that we can actually do machine ethics via theorem-proving (or model-checking).

To end on a slightly less technical note, we finish this section by addressing some interesting properties concerning interactions of certain operators in the logic presented:

$$\not\models \odot[\alpha]\varphi \rightarrow \Diamond K_\alpha[\alpha]\varphi.$$

‘It is not the case that if the agent objectively ought to do something, then it can knowingly do it.’ Our solution to Example 3 poses a counterexample, because $\langle m_2, h_1 \rangle$ is such that $\langle m_2, h_1 \rangle \models \odot[\alpha]W$, but $\langle m_2, h_1 \rangle \not\models \Diamond K_\alpha[\alpha]W$, as witnessed by the facts that $\langle m_2, h_2 \rangle \not\models [\alpha]W$ and that $\langle m_3, h_4 \rangle \not\models [\alpha]W$.

$$\not\models \odot_S[\alpha]\varphi \rightarrow \odot[\alpha]\varphi.$$

‘It is not the case that if the agent subjectively ought to do something, then it objectively ought to do it.’ Our solution to Example 2 poses a counterexample, because $\langle m_2, h_1 \rangle$ is such that $\langle m_2, h_1 \rangle \models \odot_S[\alpha]\neg G$, but $\langle m_2, h_1 \rangle \not\models \odot[\alpha]\neg G$.

$$\not\models \odot[\alpha]\varphi \rightarrow \odot_S[\alpha]\varphi.$$

‘It is not the case that if the agent objectively ought to do something, then it subjectively ought to do it.’ Our solution to Example 3 poses a counterexample, because $\langle m_2, h_1 \rangle$ is such that $\langle m_2, h_1 \rangle \models \odot[\alpha]W$, but $\langle m_2, h_1 \rangle \not\models \odot_S[\alpha]W$.

5 Conclusion

This work deals with important questions in the modeling of agency, knowledge, and obligation. Formal depictions of such concepts are likely to be useful when it comes to doing machine ethics based on deontic logic and its mechanization. The approach discussed and analyzed here is based on a stit logic of utilitarian ‘ought-to-do’ enriched with epistemic relations. We argue that to solve certain problems in the treatment of knowledge and obligations within stit –namely Horty’s puzzles– one possibility is to distinguish between objective and subjective versions of the ought-to-do modality. Moreover, we show that this possibility comes with formal advantages such as simplicity of semantics and axiomatizability.

References

1. Arkoudas, K., Bringsjord, S., Bello, P.: Toward ethical robots via mechanized deontic logic. In: AAAI Fall Symposium on Machine Ethics. pp. 17–23 (2005)
2. Belnap, N., Perloff, M., Xu, M.: Facing the future: agents and choices in our indeterminist world. Oxford University Press (2001)
3. Bringsjord, S., Arkoudas, K., Bello, P.: Toward a general logicist methodology for engineering ethically correct robots. IEEE Intelligent Systems **21**(4), 38–44 (2006)
4. Broersen, J.: Deontic epistemic stit logic distinguishing modes of mens rea. Journal of Applied Logic **9**(2), 137–152 (2011)
5. Broersen, J., Ramírez Abarca, A.I.: Formalising oughts and practical knowledge without resorting to action types. In: Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems. pp. 1877–1879. International Foundation for Autonomous Agents and Multiagent Systems (2018)
6. Horty, J.: Epistemic oughts in stit semantics (2018)
7. Horty, J., Pacuit, E.: Action types in stit semantics. Review of Symbolic Logic (2017), forthcoming
8. Horty, J.F.: Agency and Deontic Logic. Oxford University Press (2001)

9. Murakami, Y.: Utilitarian deontic logic. *AiML-2004: Advances in Modal Logic* **287** (2004)
10. Pereira, L.M., Saptawijaya, A., et al.: *Programming machine ethics*, vol. 26. Springer (2016)
11. Xu, M.: Combinations of stit with ought and know. *Journal of Philosophical Logic* **44**(6), 851–877 (2015). <https://doi.org/10.1007/s10992-015-9365-7>, <http://dx.doi.org/10.1007/s10992-015-9365-7>

A Proof of Soundness

Proposition 2. *The system Λ is sound with respect to the class of epistemic utilitarian bi-valued BT-models.*

Proof. Let $\mathcal{M} = \langle T, \sqsubset, \mathbf{Choice}, \{\sim_\alpha\}_{\alpha \in Ags}, \mathbf{Value}_O, \mathbf{Value}_S \rangle$ be an epistemic utilitarian bi-valued BT-frame. Let \mathcal{V} be any pertinent valuation function. The **S5** axioms for \square , $[\alpha]$, K_α , as well as axioms (A1), (A2), (A3), (A4), (A5), (A6), (Oic), (IA), (OAC), and (Unif – H) are shown to be semantic validities straightforwardly. Since they involve some novelty, we include the detailed proofs for (s.N), (s.Oic), and (Cl) here.

- To see that $\mathcal{M} \models (s.N)$, take $\langle m_*, h_* \rangle$ such that $\langle m_*, h_* \rangle \models K_\alpha \square \varphi$. Let m be such that $m_* \sim_\alpha m$ (which means that there exist $j \in H_{m_*}$, $j' \in H_m$ such that $\langle m_*, j \rangle \sim_\alpha \langle m, j' \rangle$), and let $L \in \mathbf{Choice}_\alpha^{m_*}$. Condition (Unif – H) ensures that there exists $h'_* \in H_m$ such that $\langle m_*, h_* \rangle \sim_\alpha \langle m, h'_* \rangle$. Our assumption that $\langle m_*, h_* \rangle \models K_\alpha \square \varphi$ implies then that $\langle m, h'_* \rangle \models \square \varphi$. Therefore, for any $h \in [L]_\alpha^m$, the fact that $h \in H_m$ yields that $\langle m, h \rangle \models \varphi$. Therefore, for every $L \in \mathbf{Choice}_\alpha^{m_*}$ and every m such that $m_* \sim_\alpha m$, we have that $[L]_\alpha^m \subseteq |\varphi|^m$, which vacuously implies that $\langle m_*, h_* \rangle \models \odot_S[\alpha]\varphi$.
- To see that $\mathcal{M} \models (s.Oic)$, take $\langle m_*, h_* \rangle$ such that $\langle m_*, h_* \rangle \models \odot_S[\alpha]\varphi$. This implies that there exists $L_* \subseteq H_{m_*}$ such that $[L_*]_\alpha^{m'} \subseteq |\varphi|^{m'}$ for every $m' \in T$ such that $m_* \sim_\alpha m'$. Since \sim_α is reflexive, it is clear that $[L_*]_\alpha^{m_*} \subseteq |\varphi|^{m_*}$. Now, take $h_0 \in L_*$. Let $\langle m, h \rangle$ be such that $\langle m_*, h_0 \rangle \sim_\alpha \langle m, h \rangle$. From the definition of the epistemic cluster set, we have that $h \in [L_*]_\alpha^m$, so that $[L_*]_\alpha^m \subseteq |\varphi|^m$ and thus $\langle m, h \rangle \models \varphi$. Therefore, we have that $h_0 \in H_{m_*}$ is such that for every $\langle m, h \rangle$ with $\langle m_*, h_0 \rangle \sim_\alpha \langle m, h \rangle$, $\langle m, h \rangle \models \varphi$. This means that $\langle m_*, h_0 \rangle \models K_\alpha \varphi$. Therefore, we have that $\langle m_*, h_* \rangle \models \Diamond K_\alpha \varphi$.
- To see that $\mathcal{M} \models (Cl)$, take $\langle m_*, h_* \rangle$ such that $\langle m_*, h_* \rangle \models \odot_S[\alpha]\varphi$. Let $\langle m, j \rangle$ be such that $\langle m_*, h_* \rangle \sim_\alpha \langle m, j \rangle$. Take $h \in H_m$. We want to show that for every $L \in \mathbf{Choice}_\alpha^m$ such that $[L]_\alpha^{m'} \not\subseteq |\varphi|^{m'}$ (for some m' such that $m \sim_\alpha m'$), there exists $L' \in \mathbf{Choice}_\alpha^m$ such that $L \prec_s L'$ and if $L'' = L'$ or $L' \preceq_s L''$, then for every m'' such that $m \sim_\alpha m''$, we have that $[L'']_\alpha^{m''} \subseteq |\varphi|^{m''}$. Take $L \in \mathbf{Choice}_\alpha^m$ such that there exists $m' \in T$ with $m \sim_\alpha m'$ and $[L]_\alpha^{m'} \not\subseteq |\varphi|^{m'}$. Let N_L be an action in $\mathbf{Choice}_\alpha^{m_*}$ such that $N_L \subseteq [L]_\alpha^{m_*}$. We know that such an action exists, in virtue of (Unif – H) and (OAC). Notice that transitivity of \sim_α entails that $[N_L]_\alpha^o = [L]_\alpha^o$ for any moment o , so we have that $[N_L]_\alpha^{m'} \not\subseteq |\varphi|^{m'}$. Since $\langle m_*, h_* \rangle \models \odot_S[\alpha]\varphi$, we get that there must exist $N \in \mathbf{Choice}_\alpha^{m_*}$ such that $N_L \prec_s N$ and if $N' = N$ or $N \preceq_s N'$, then for every m'' such that $m_* \sim_\alpha m''$, $[N']_\alpha^{m''} \subseteq |\varphi|^{m''}$. Now, let L_N be an action in \mathbf{Choice}_α^m such that $L_N \subseteq [N]_\alpha^m$ (which implies that $[L_N]_\alpha^o = [N]_\alpha^o$ for any moment o). We claim that $L \prec_s L_N$, and show our claim with the following argument. Let m'' be a moment such that $m \sim_\alpha m''$, and let $S \in \mathbf{State}_\alpha^{m''}$. On one hand, we have that $\star [L]_\alpha^{m''} \cap S = [N_L]_\alpha^{m''} \cap S \leq_s [N]_\alpha^{m''} \cap S = [L_N]_\alpha^{m''} \cap S$. On the other hand, we know that there exists a moment m''' such that $m_* \sim_\alpha m'''$ and a state $S_0 \in \mathbf{State}_\alpha^{m'''}$ such that $[N]_\alpha^{m'''} \cap S_0 \not\leq_s [N_L]_\alpha^{m'''} \cap S_0$. Therefore, we have that $(\star) [L_N]_\alpha^{m''} \cap S_0 =$

$[N]_\alpha^{m'''} \cap S_0 \not\leq_s [N_L]_\alpha^{m'''} \cap S_0 = [L]_\alpha^{m'''} \cap S_0$. Together, (\star) and $(\star\star)$ entail that $L \prec_s L_N$. Now, let $L'' \in \mathbf{Choice}_\alpha^m$ such that $L'' = L_N$ or $L_N \preceq_s L''$. If $L'' = L_N$, then for every m'' such that $m \sim_\alpha m''$, $[L'']_\alpha^{m''} = [N]_\alpha^{m''} \subseteq |\varphi|^{m''}$. If $L_N \prec_s L''$, then an argument similar to the one we used above to show that our claim was true renders that an action $N_{L''} \in \mathbf{Choice}_\alpha^{m_*}$ with $N_{L''} \subseteq [L'']_\alpha^{m_*}$ is such that $N \preceq_s N_{L''}$, so that $[L'']_\alpha^{m''} = [N_{L''}]_\alpha^{m''} \subseteq |\varphi|^{m''}$. With this, we have shown that $\langle m, h \rangle \models \odot_S [\alpha] \varphi$. Since this happens for every $h \in H_m$, we have that $\langle m, j \rangle \models \Box \odot_S [\alpha] \varphi$. Since this happens for any $\langle m, j \rangle$ such that $\langle m_*, h_* \rangle \sim_\alpha \langle m, j \rangle$, then we get that $\langle m_*, h_* \rangle \models K_\alpha \Box \odot_S [\alpha] \varphi$.

– *Modus Ponens*, Substitution, and Necessitation for \Box preserve validity.

Therefore, we have shown that the system Λ is sound with respect to the class of epistemic bi-valued BT-models.

B Proof of Completeness

Definition 5 (Kripke-estit frames).

A tuple $\mathcal{F} = \langle W, R_\Box, \mathbf{Choice}, \{\approx_\alpha\}_{\alpha \in \text{Ags}}, \text{Value}_O, \text{Value}_S \rangle$ is called a Kripke-estit frame iff W is a set of possible worlds and

1. R_\Box is an equivalence relation over W . For $w \in W$, the class of w under R_\Box is denoted by \bar{w} .
2. \mathbf{Choice} is a function that assigns to each $\alpha \in \text{Ags}$ and each \Box -class \bar{w} (with $w \in W$) a partition $\mathbf{Choice}_\alpha^{\bar{w}}$ of \bar{w} given by an equivalence relation which we will denote by $R_\alpha^{\bar{w}}$. \mathbf{Choice} must satisfy the following constraint:
 - $(\mathbf{IA})_K$ For $w \in W$, we have that each function $s : \text{Ags} \rightarrow \mathcal{P}(\bar{w})$ that maps α to a member of $\mathbf{Choice}_\alpha^{\bar{w}}$ is such that $\bigcap_{\alpha \in \text{Ags}} s(\alpha) \neq \emptyset$ (the set of all functions s that map α to a member of $\mathbf{Choice}_\alpha^{\bar{w}}$ is denoted by \mathbf{Select}^m).

For $w \in W$ and $v \in \bar{w}$, the class of v in the partition $\mathbf{Choice}_\alpha^{\bar{w}}$ is denoted by $\mathbf{Choice}_\alpha^{\bar{w}}(v)$. For $w \in W$ and $\beta \in \text{Ags}$, we define $\mathbf{State}_\beta^{\bar{w}} = \{S \subseteq \bar{w}; S = \bigcap_{\alpha \in \text{Ags} - \{\beta\}} s(\alpha), \text{ where } s \in \mathbf{Select}^m\}$.
3. For each $\alpha \in \text{Ags}$, \approx_α is an (epistemic) equivalence relation on W that satisfies the following constraints:
 - $(\mathbf{OAC})_K$ For $w \in W$ such that $v \in \bar{w}$, if $v \approx_\alpha u$, then $v' \approx_\alpha u$ for every $v' \in \mathbf{Choice}_\alpha^{\bar{w}}(v)$.
 - $(\mathbf{Unif} - \mathbf{H})_K$ Let $w_1, w_2 \in W$ such that there exist $v \in \bar{w}_1$ and $u \in \bar{w}_2$ with $v \approx_\alpha u$. Then for every $v' \in \bar{w}_1$, there exists $u' \in \bar{w}_2$ such that $v' \approx_\alpha u'$.

For each $v \in W$ such that $w \sim_\alpha v$, we define the epistemic cluster set of L at \bar{v} by $\llbracket L \rrbracket_\alpha^{\bar{v}} := \{u \in \bar{v}; \exists o \in L \text{ such that } o \approx_\alpha u\}$.
4. Value_O and Value_S are (utilitarian) deontic functions that assign to each $w \in W$ a value in \mathbb{R} . We use them to define objective and subjective orderings just as for BT-models.

Kripke-estit frames allow us to evaluate the formulas of \mathcal{L}_{KO} in an analogous fashion to the semantics provided for BT-frames. Therefore, we use the name ‘Kripke-estit models’ for the structures that result from adding a valuation function to Kripke-estit

frames, where the semantics for the formulas of \mathcal{L}_{KO} is defined so as to mirror Definition 3. Kripke-estit models can be used for constructing epistemic utilitarian bi-valued BT-models such that both satisfy the same formulas of \mathcal{L}_{KO} :

Definition 6 (Associated BT-frame).

Let $\mathcal{F} = \langle W, R_\square, \text{Choice}, \{\approx_\alpha\}_{\alpha \in \text{Ags}}, \text{Value}_O, \text{Value}_S \rangle$ be a Kripke-estit frame. We define the tuple $\langle T_W, \sqsubset, \text{Choice}, \{\sim_\alpha\}_{\alpha \in \text{Ags}}, \text{Value}_O, \text{Value}_S \rangle$ will be called the epistemic utilitarian bi-valued BT-frame associated to \mathcal{F} such that

1. $T_W := W \cup \{\bar{w}; w \in W\} \cup \{W\}$ and \sqsubset is a relation on T_W such that \sqsubset is defined as the transitive closure of the union $\{(\bar{w}, v); w \in W \text{ and } v \in \bar{w}\} \cup \{(W, \bar{w}); w \in W\}$. Notice that \sqsubset is a strict partial order on T_W that satisfies “no backward branching” straightforwardly. Since the tuple $\langle T_W, \sqsubset \rangle$ is thus a tree, we can call the maximal \sqsubset -chains in T_W histories. We observe that the definition of \sqsubset yields that there is a bijective correspondence between the set of worlds in W and the histories of T_W . If for each $v \in W$, we set h_v to be the history $\{W, \bar{v}, v\}$, then we have that for any world $o \in W$, $o \in h_v$ iff $o = v$. Therefore, we will conveniently identify these histories using the worlds at their terminal nodes. In this line of thought, we have that for each $w \in W$, the set $H_{\bar{w}}$ of histories running through \bar{w} is equal to the set $\{h_v; v \in \bar{w}\}$ —since $\bar{w} \in h_v$ iff $v \in \bar{w}$. Similarly, we have that the set H of all histories is the same as $\{h_v; v \in W\}$.
2. **Choice** is an agentive choice function for the tree of the above item. In order to define such a function, we use the partitions $\text{Choice}_\alpha^{\bar{w}}$ (where $w \in W$ and $\alpha \in \text{Ags}$) in the following manner: for each $B \in \mathcal{P}(W)$, we will write B^T to denote the set $\{h_v; v \in B\}$; this way, we define a function **Choice** on $\text{Ags} \times T_W$ with extension given by $\text{Choice}(\alpha, v) = \{\{h_v\}\}$ for each $\alpha \in \text{Ags}$ and $v \in W$, $\text{Choice}(\alpha, \bar{w}) = \{C_\alpha^T; C_\alpha \in \text{Choice}_\alpha^{\bar{w}}\}$ for each $w \in W$ and each $\alpha \in \text{Ags}$, and $\text{Choice}(\alpha, W) = \{H\}$ for each $\alpha \in \text{Ags}$. To keep notation consistent, we will denote sets of the form $\text{Choice}(\alpha, \bar{w})$ by $\text{Choice}_\alpha^{\bar{w}}$, so that the choice-class of a given h_v in $\text{Choice}_\alpha^{\bar{w}}$ is denoted by $\text{Choice}_\alpha^{\bar{w}}(h_v)$, as expected.¹¹ It is important to mention that this definition yields that for every $S \in \text{State}_\alpha^{\bar{w}}$, $S^T \in \text{State}_\alpha^{\bar{w}}$, and that for each $U \in \text{State}_\alpha^{\bar{w}}$, there exists $V \in \text{State}_\alpha^{\bar{w}}$ such that $U = V^T$.
3. For each $\alpha \in \text{Ags}$, we define an epistemic relation \sim_α on T_W using the set $\{\approx_\alpha\}_{\alpha \in \text{Ags}}$. For a given agent $\alpha \in \text{Ags}$, we define $\sim_\alpha = \{(\langle \bar{w}, h_v \rangle, \langle \bar{w}', h_{v'} \rangle); w, w' \in W \text{ and } v \approx_\alpha v'\} \cup \{(\langle z, h_z \rangle, \langle z, h_z \rangle); z \in W\} \cup \{(\langle W, h_v \rangle, \langle W, h_{v'} \rangle); v, v' \in W\}$. We observe, then, that this definition entails that \sim_α is an equivalence relation for every $\alpha \in \text{Ags}$ and that, for a given $w \in W$ and $L \in \text{Choice}_\alpha^{\bar{w}}$, $v \in \llbracket L \rrbracket_\alpha^{\bar{w}}$ iff $h_v \in [L^T]_\alpha^{\bar{w}}$.
4. **Value_O** and **Value_S** are two utilitarian deontic functions, defined such that for each $h_v \in H$, **Value_O**(h_v) = **Value_O**(v) and **Value_S**(h_v) = **Value_S**(v). Endowed with these deontic functions, we can define objective and subjective dominance orderings for this structure accordingly.

Proposition 3. Let $\mathcal{F} = \langle W, R_\square, \text{Choice}, \{\approx_\alpha\}_{\alpha \in \text{Ags}}, \text{Value}_O, \text{Value}_S \rangle$ be a Kripke-estit frame. The tuple $\langle T_W, \sqsubset, \text{Choice}, \{\sim_\alpha\}_{\alpha \in \text{Ags}}, \text{Value}_O, \text{Value}_S \rangle$ is an epistemic utilitarian bi-valued BT-frame.

¹¹ Notice that this implies that, for two worlds v, v' in \bar{w} , $v R_\alpha^{\bar{w}} v'$ iff $h_v \in \text{Choice}_\alpha^{\bar{w}}(h_{v'})$.

Proof. It amounts to showing that the tuple, as defined, validates the semantic conditions (NC), (IA), (OAC), and (Unif – H).

- (NC) is vacuously validated in moment W . It is validated in moments of the form \bar{w} ($w \in W$), since two different histories never intersect in a moment later than \bar{w} . Finally, it is also validated in moments of the form v such that $v \in W$ (since there are no moments above v).
- For (IA), we reason by cases:
 - a) At moment W , (IA) is validated straightforwardly, since $\mathbf{Choice}(\alpha, W) = \{H\}$ for each $\alpha \in \mathcal{Ags}$.
 - b) For a fixed moment of the form \bar{w} (with $w \in W$), let s be a function that assigns to each agent α a member of $\mathbf{Choice}_{\alpha}^{\bar{w}} = \{(C_{\alpha})^T; C_{\alpha} \in \mathbf{Choice}_{\alpha}^{\bar{w}}\}$. Let $s_k : \mathcal{Ags} \rightarrow \bigcup_{\alpha \in \mathcal{Ags}} \mathbf{Choice}_{\alpha}^{\bar{w}}$ be a function such that $s_k(\alpha) = C_{\alpha}$ iff $s(\alpha) = (C_{\alpha})^T$. Since \mathcal{M} validates the constraint (IA) $_K$, then we have that $\bigcap_{\alpha \in \mathcal{Ags}} s_k(\alpha) \neq \emptyset$. Take $v \in \bigcap_{\alpha \in \mathcal{Ags}} s_k(\alpha)$. Then $v \in C_{\alpha}$ for every $\alpha \in \mathcal{Ags}$. This implies that $h_v \in (C_{\alpha})^T$ for every $\alpha \in \mathcal{Ags}$, so that $\bigcap_{\alpha \in \mathcal{Ags}} s(\alpha) \neq \emptyset$.
 - c) At moments of the form v such that $v \in W$, if s is a function that assigns to each agent α a member of $\mathbf{Choice}(v, \alpha)$, we have that s must be constant and $\bigcap_{\alpha \in \mathcal{Ags}} s(\alpha) = \{h_v\}$.
- For (OAC), again we reason by cases:
 - a) Assume that $\langle \bar{w}, h_v \rangle \sim_{\alpha} \langle \bar{w}', h_{v'} \rangle$ (for $w, w' \in W$). This means that $v \approx_{\alpha} v'$. We want to show that for every $h_u \in \mathbf{Choice}_{\alpha}^{\bar{w}}$ such that $h_u \in \mathbf{Choice}_{\alpha}^{\bar{w}}(h_v)$, $\langle \bar{w}, h_u \rangle \sim_{\alpha} \langle \bar{w}', h_{v'} \rangle$. Therefore, let $h_u \in \mathbf{Choice}_{\alpha}^{\bar{w}}(h_v)$. By definition, we have that this means that $u \in \mathbf{Choice}_{\alpha}^{\bar{w}}(v)$. Since \mathcal{M} validates the constraint (OAC) $_K$, this last fact implies, with $v \approx_{\alpha} v'$, that $u \approx_{\alpha} v'$, which in turn yields that $\langle \bar{w}, h_u \rangle \sim_{\alpha} \langle \bar{w}', h_{v'} \rangle$.
 - b) For situations with moments of the form v such that $v \in W$, we have that (OAC) is met straightforwardly, since for h_v , the choice-cell in $\mathbf{Choice}(\alpha, v)$ to which h_v belongs is just $\{h_v\}$.
 - c) For situations with moment W , (OAC) is also met straightforwardly, since for every $\alpha \in \mathcal{Ags}$, \sim_{α} is defined such that $\langle W, h_v \rangle \sim_{\alpha} \langle W, h_{v'} \rangle$ for every pair of histories $h_v, h_{v'}$ in H .
- For (Unif – H), again we reason by cases:
 - a) Assume that $\langle \bar{w}, h_v \rangle \sim_{\alpha} \langle \bar{w}', h_{v'} \rangle$ (for $w, w' \in W$). This means that $v \in \bar{w}$, $v' \in \bar{w}'$, and $v \approx_{\alpha} v'$. Let $h_z \in H_{\bar{w}}$ (which means that $z \in \bar{w}$). We want to show that there exists $h \in H_{\bar{w}'}$ such that $\langle \bar{w}, h_z \rangle \sim_{\alpha} \langle \bar{w}', h \rangle$. Condition (Unif – H) $_K$ gives us that there exists $z' \in \bar{w}'$ such that $z \approx_{\alpha} z'$, which by definition of \sim_{α} means that $\langle \bar{w}, h_z \rangle \sim_{\alpha} \langle \bar{w}', h_{z'} \rangle$.
 - b) For situations with moments of the form v such that $v \in W$, we have that $\langle v, h_v \rangle$ is \sim_{α} -related only to itself, so that (Unif – H) is met straightforwardly.
 - c) For situations with moment W , (Unif – H) is also met straightforwardly, since $\langle W, h_v \rangle \sim_{\alpha} \langle W, h_{v'} \rangle$ for every $v, v' \in W$.

Let \mathcal{M} be a Kripke-estit model with valuation function \mathcal{V} . The frame upon which \mathcal{M} is based has an associated BT-frame. If to the tuple of this BT-frame we add a valuation

function \mathcal{V}^T such that $\mathcal{V}^T(p) = \{\langle \bar{w}, h_w \rangle; w \in \mathcal{V}(p)\}$, we call the resulting model the epistemic utilitarian bi-valued BT-model associated to \mathcal{M} .

Proposition 4. *Let \mathcal{M} be a Kripke-estit model, and let \mathcal{M}^t denote its associated BT-model. For a formula φ of \mathcal{L}_{KO} and $w \in W$, we have that $\mathcal{M}, w \models \varphi$ iff $\mathcal{M}^t, \langle \bar{w}, h_w \rangle \models \varphi$.*

Proof. **Claim 1.** For a given Kripke-estit model \mathcal{M} and a fixed $w \in W^A$, we have that for every $L, N \in \text{Choice}_{\alpha}^{\bar{w}}$,

- $L \preceq N$ iff $L^T \preceq N^T$ and $L \prec N$ iff $L^T \prec N^T$.
- $L \preceq_s N$ iff $L^T \preceq_s N^T$ and $L \prec_s N$ iff $L^T \prec_s N^T$.
- $L \in \text{Optimal}_{\alpha}^{\bar{w}}$ iff $L^T \in \text{Optimal}_{\alpha}^{\bar{w}}$, and $L \in \mathbf{S} - \text{Optimal}_{\alpha}^{\bar{w}}$ iff $L^T \in \mathbf{S} - \text{Optimal}_{\alpha}^{\bar{w}}$.

We will prove **Claim 1** only for the strict orderings, since these proofs include the arguments needed to show that the statements for \preceq and \preceq_s also hold.

- (\Rightarrow) We assume that $L \prec N$. Let $U \in \text{State}_{\alpha}^{\bar{w}}$. We know that $U = V^T$ for some $V \in \text{State}_{\alpha}^{\bar{w}}$. Therefore, we have that $L \cap V \leq N \cap V$. But notice that, by definition, we have that for each $v \in \bar{w}$, $\text{Value}_{\mathcal{O}}(h_v) = \text{Value}_{\mathcal{O}}(v)$, so the fact that $L \cap V \leq N \cap V$ implies that $L^T \cap U \leq N^T \cap U$. Now, our assumption also yields that there should exist $S_0 \in \text{State}_{\alpha}^{\bar{w}}$ such that $N \cap S_0 \not\leq L \cap S_0$. This straightforwardly implies that $N^T \cap S_0^T \not\leq L^T \cap S_0^T$, so that indeed we have that $L^T \prec N^T$.
 (\Leftarrow) We assume that $L^T \prec N^T$. Let $S \in \text{State}_{\alpha}^{\bar{w}}$. We know that $L^T \cap S^T \leq N^T \cap S^T$, so that $L \cap S \leq N \cap S$. Similarly, here our assumption yields that there should exist $U_0 \in \text{State}_{\alpha}^{\bar{w}}$ such that $N^T \cap U_0 \not\leq L^T \cap U_0$. But we have that $U_0 = V_0^T$ for some $V_0 \in \text{State}_{\alpha}^{\bar{w}}$, so that the definition of $\text{Value}_{\mathcal{O}}$ implies that $N \cap V_0 \not\leq L \cap V_0$. With this, we have shown that $L \prec N$.
- (\Rightarrow) We assume that $L \prec_s N$. Let w' be such that $w \approx_{\alpha} w'$, and let $U \in \text{State}_{\alpha}^{\bar{w}'}$. We know that $U = V^T$ for some $V \in \text{State}_{\alpha}^{\bar{w}'}$. Therefore, we have that $\llbracket L \rrbracket_{\alpha}^{\bar{w}'} \cap V \leq_s \llbracket N \rrbracket_{\alpha}^{\bar{w}'} \cap V$. We have that for each $w \in W$, $\text{Value}_S(h_w) = \text{Value}_S(w)$, and that for each pair v_1, v_2 such that $v_1 \in \llbracket L \rrbracket_{\alpha}^{\bar{w}'} \cap V$ and $v_2 \in \llbracket N \rrbracket_{\alpha}^{\bar{w}'} \cap V$, then $h_{v_1} \in \llbracket L^T \rrbracket_{\alpha}^{\bar{w}'} \cap V^T$ and $h_{v_2} \in \llbracket N^T \rrbracket_{\alpha}^{\bar{w}'} \cap V^T$, so that the fact that $\llbracket L \rrbracket_{\alpha}^{\bar{w}'} \cap V \leq_s \llbracket N \rrbracket_{\alpha}^{\bar{w}'} \cap V$ then implies that $\llbracket L^T \rrbracket_{\alpha}^{\bar{w}'} \cap U \leq_s \llbracket N^T \rrbracket_{\alpha}^{\bar{w}'} \cap U$. Now, our assumption also yields that there should exist w_* such that $w \approx_{\alpha} w_*$ and $S_0 \in \text{State}_{\alpha}^{\bar{w}_*}$ such that $\llbracket N \rrbracket_{\alpha}^{\bar{w}_*} \cap S_0^T \not\leq_s \llbracket L \rrbracket_{\alpha}^{\bar{w}_*} \cap S_0^T$. Analogously, this implies that $\llbracket N^T \rrbracket_{\alpha}^{\bar{w}_*} \cap S_0 \not\leq_s \llbracket L^T \rrbracket_{\alpha}^{\bar{w}_*} \cap S_0$, so that indeed we have that $L^T \prec_s N^T$.
 (\Leftarrow) We assume that $L^T \prec_s N^T$. Let w' be such that $w \approx_{\alpha} w'$, and let $S \in \text{State}_{\alpha}^{\bar{w}'}$. We know that $\llbracket L^T \rrbracket_{\alpha}^{\bar{w}'} \cap S^T \leq_s \llbracket N^T \rrbracket_{\alpha}^{\bar{w}'} \cap S^T$, so that an argument similar to the one displayed for the above item renders that $\llbracket L \rrbracket_{\alpha}^{\bar{w}'} \cap S \leq_s \llbracket N \rrbracket_{\alpha}^{\bar{w}'} \cap S$. On the other hand, there should exist w_* such that $w \approx_{\alpha} w_*$ and $U_0 \in \text{State}_{\alpha}^{\bar{w}_*}$ such that $N^T \cap U_0 \not\leq_s L^T \cap U_0$. But we have that $U_0 = V_0^T$ for some $V_0 \in \text{State}_{\alpha}^{\bar{w}_*}$, so that, in an analogous fashion to what we did above, we get that $N \cap V_0 \not\leq_s L \cap V_0$. With this, we have shown that $L \prec_s N$.
- Straightforward from the above items.

With that we conclude the proof of **Claim 1**. For the proof of the main statement, we proceed by induction on φ . For the base case, we have that, for a propositional letter p and an arbitrary $w \in W$, $\mathcal{M}, w \Vdash p$ iff $w \in \mathcal{V}(p)$ iff $\langle \bar{w}, h_w \rangle \in \mathcal{V}^T(p)$ iff $\mathcal{M}^T, \langle \bar{w}, h_w \rangle \models p$. The cases with the boolean connectives are standard, so let us deal with the modal operators. Let $w \in W$ and $\alpha \in \text{Ags}$.

- (“ \Box ”)
We have that $\mathcal{M}, w \Vdash \Box\varphi$ iff for every $v \in \bar{w}$, $\mathcal{M}_w, v \Vdash \varphi$, which by induction hypothesis happens iff $\mathcal{M}^T, \langle \bar{v}, h_v \rangle \models \varphi$ for every $v \in \bar{w}$, which happens iff $\mathcal{M}^T, \langle \bar{w}, h_w \rangle \models \Box\varphi$, since we have that $h_v \in H_{\bar{w}}$ iff $v \in \bar{w}$.
- (“ $[\alpha]$ ”)
We have that $\mathcal{M}, w \Vdash [\alpha]\varphi$ iff for every $v \in W$ such that $w R_{\alpha}^{\bar{w}} v$, $\mathcal{M}, v \Vdash \varphi$, which by induction hypothesis happens iff $\mathcal{M}^T, \langle \bar{w}, h_v \rangle \models \varphi$ for every $h_v \in \mathbf{Choice}_{\alpha}^{\bar{w}}(h_w)$, which in turn happens iff $\mathcal{M}^T, \langle \bar{w}, h_w \rangle \models [\alpha]\varphi$.
- (“ K_{α} ”)
We have that $\mathcal{M}, w \Vdash K_{\alpha}\varphi$ iff for every $v \in W$ such that $w \approx_{\alpha} v$, $\mathcal{M}, v \Vdash \varphi$, which by induction hypothesis occurs iff $\mathcal{M}^T, \langle \bar{v}, h_v \rangle \models \varphi$ for every $h_v \in H$ such that $\langle \bar{w}, h_w \rangle \sim_{\alpha} \langle \bar{v}, h_v \rangle$, which happens iff $\mathcal{M}^T, \langle \bar{w}, h_w \rangle \models K_{\alpha}\varphi$.
- (“ $\odot[\alpha]$ ”)
(\Rightarrow) We assume that $\mathcal{M}, w \Vdash \odot[\alpha]\varphi$. Then for every $L \in \mathbf{Choice}_{\alpha}^{\bar{w}}$ such that $\mathcal{M}, v_L \not\models \varphi$ for some $v_L \in L$, there exists $L' \in \mathbf{Choice}_{\alpha}^{\bar{w}}$ such that $L \prec L'$ and if $L'' = L'$ or $L' \preceq L''$, then $\mathcal{M}, v \Vdash \varphi$ for every $v \in L''$. Let $N \in \mathbf{Choice}_{\alpha}^{\bar{w}}$ such that $\mathcal{M}^T, \langle \bar{w}, h_z \rangle \not\models \varphi$ for some $h_z \in N$. By induction hypothesis, we get that $\mathcal{M}, z \not\models \varphi$. The cell N is L_0^T for some $L_0 \in \mathbf{Choice}_{\alpha}^{\bar{w}}$. Since $z \in L_0$ and $\mathcal{M}, z \not\models \varphi$, our assumption entails that there exists $L_1 \in \mathbf{Choice}_{\alpha}^{\bar{w}}$ such that $L_0 \prec L_1$ and if $L'' = L_1$ or $L_1 \preceq L''$ then $\mathcal{M}, v \Vdash \varphi$ for every $v \in L''$. We claim that L_1^T is the choice cell at moment \bar{w} that will witness to the fact that $\mathcal{M}^T, \langle \bar{w}, h_w \rangle \models \odot[\alpha]\varphi$. In order to show this, we first notice that the first item of **Claim 1** renders that $L_0^T \prec L_1^T$. Now, we also have that for every $h_v \in L_1^T$, $v \in L_1$ (observe also that $v \in \bar{w}$ and therefore $\mathcal{M}, v \Vdash \varphi$; in this way, we have that for every $h_v \in L_1^T$, the induction hypothesis gives that $\mathcal{M}^T, \langle \bar{w}, h_v \rangle \models \varphi$. Finally, let $V^T \in \mathbf{Choice}_{\alpha}^{\bar{w}}$ such that $L_1^T \preceq V^T$. By the first item of **Claim 1**, this last condition implies that $L_1 \preceq V$. Take $h_v \in V^T$. We know that $v \in V$, so that the fact that $L_1 \preceq V$ implies with our assumption that $\mathcal{M}, v \Vdash \varphi$; therefore, the induction hypothesis gives that $\mathcal{M}^T, \langle \bar{w}, h_v \rangle \models \varphi$. With this, we have shown that our claim is true and that $\mathcal{M}^T, \langle \bar{w}, h_w \rangle \models \odot[\alpha]\varphi$.
(\Leftarrow) We assume that $\mathcal{M}^T, \langle \bar{w}, h_w \rangle \models \odot[\alpha]\varphi$. Then for every $L^T \in \mathbf{Choice}_{\alpha}^{\bar{w}}$ such that $\mathcal{M}^T, \langle \bar{w}, h_z \rangle \not\models \varphi$ for some $h_z \in L^T$, there exists $L_1^T \in \mathbf{Choice}_{\alpha}^{\bar{w}}$ such that $L^T \prec_1^T$ and if $U = L_1^T$ or $L^T \preceq U$, then $\mathcal{M}^T, \langle \bar{w}, h_u \rangle \models \varphi$ for every $h_u \in U$. Let $N \in \mathbf{Choice}_{\alpha}^{\bar{w}}$ such that $\mathcal{M}, v \not\models \varphi$ for some $v \in N$. By induction hypothesis, we get that $\mathcal{M}_w^T, \langle \bar{w}, h_v \rangle \not\models \varphi$. We have that $N^T \in \mathbf{Choice}_{\alpha}^{\bar{w}}$. Therefore, our assumption entails that there exists $L_1^T \in \mathbf{Choice}_{\alpha}^{\bar{w}}$ such that $N^T \prec L_1^T$ and if $U = L_1^T$ or $L_1^T \preceq U$, then $\mathcal{M}^T, \langle \bar{w}, h_u \rangle \models \varphi$ for every $h_u \in U$. Here, we claim that L_1 is the choice cell that will witness to the fact that $\mathcal{M}, w \Vdash \odot[\alpha]\varphi$. In order to show this, we first notice that the first item of **Claim 1** renders that $N \prec L_1$. Next, for every $v \in L_1$, $h_v \in L_1^T$ and therefore $\mathcal{M}^T, \langle \bar{w}, h_v \rangle \models \varphi$; in this way, we have

that for every $v \in L_1$, the induction hypothesis gives that $\mathcal{M}, v \models \varphi$. Finally, let $L \in \text{Choice}_{\alpha}^{\bar{w}}$ such that $L_1 \preceq L$. By the first item of **Claim 1**, this last condition implies that $L_1^T \preceq L^T$. Take $v \in L$. We know that $h_v \in L^T$, so that the fact that $L_1^T \preceq L^T$ implies with our assumption that $\mathcal{M}^T, \langle \bar{w}, h_v \rangle \models \varphi$; therefore, the induction hypothesis gives that $\mathcal{M}, v \models \varphi$. With this, we have shown that our claim is true and that $\mathcal{M}_w, w \models \odot[\alpha]\varphi$.

– (“ $\odot_S[\alpha]$ ”)

(\Rightarrow) We assume that $\mathcal{M}, w \models \odot_S[\alpha]\varphi$. Then for every $L \in \text{Choice}_{\alpha}^{\bar{w}}$ such that $\mathcal{M}, v_L \not\models \varphi$ for some $v_L \in \llbracket L \rrbracket_{\alpha}^{\bar{w}}$ (with $w \approx_{\alpha} w'$), there exists $L' \in \text{Choice}_{\alpha}^{\bar{w}}$ such that $L \prec_s L'$ and if $L'' = L'$ or $L' \preceq_s L''$, then for every \bar{w}'' such that $w \approx_{\alpha} w''$, $\mathcal{M}, v \models \varphi$ for every $v \in \llbracket L'' \rrbracket_{\alpha}^{\bar{w}''}$. Let $N \in \text{Choice}_{\alpha}^{\bar{w}}$ such that $\mathcal{M}_W^T, \langle \bar{w}', h_z \rangle \not\models \varphi$ for some $h_z \in [N]_{\alpha}^{\bar{w}'}$ (with $\bar{w} \sim_{\alpha} \bar{w}'$). By induction hypothesis, we get that $\mathcal{M}, z \not\models \varphi$. The cell N is L_0^T for some $L_0 \in \text{Choice}_{\alpha}^{\bar{w}}$. Moreover, our definition of \sim_{α} in terms of \approx_{α} entails that $z \in \llbracket L_0 \rrbracket_{\alpha}^{\bar{w}'}$, so that the fact that $\mathcal{M}, z \not\models \varphi$ entails with our assumption that there exists $L_1 \in \text{Choice}_{\alpha}^{\bar{w}}$ such that $L_0 \prec_s L_1$ and if $L'' = L_1$ or $L_1 \preceq_s L''$ then for every w'' such that $w \approx_{\alpha} w''$, $\mathcal{M}, v \models \varphi$ for every $v \in \llbracket L'' \rrbracket_{\alpha}^{\bar{w}''}$. We claim that L_1^T is the choice cell at moment \bar{w} that will witness to the fact that $\mathcal{M}^T, \langle \bar{w}, h_w \rangle \models \odot_S[\alpha]\varphi$. In order to show this, we first notice that the second item of **Claim 1** renders that $L_0^T \prec_s L_1^T$. We also have that for every \bar{w}'' such that $\bar{w} \sim_{\alpha} \bar{w}''$ and $h_v \in [L_1^T]_{\alpha}^{\bar{w}'}$, $v \in \llbracket L_1 \rrbracket_{\alpha}^{\bar{w}''}$ and therefore $\mathcal{M}, v \models \varphi$; in this way, we have that for every \bar{w}'' such that $\bar{w} \sim_{\alpha} \bar{w}''$ and every $h_v \in [L_1^T]_{\alpha}^{\bar{w}'}$, the induction hypothesis gives that $\mathcal{M}^T, \langle \bar{w}'', h_v \rangle \models \varphi$. Finally, let $V^T \in \text{Choice}_{\alpha}^{\bar{w}}$ such that $L_1^T \preceq_s V^T$. By the second item of **Claim 1**, this last condition implies that $L_1 \preceq_s V$. Take \bar{w}'' such that $\bar{w} \sim_{\alpha} \bar{w}''$, and take $h_v \in [V^T]_{\alpha}^{\bar{w}'}$. We know that $v \in \llbracket V \rrbracket_{\alpha}^{\bar{w}''}$, so that the fact that $L_1 \preceq_s V$ implies with our assumption that $\mathcal{M}, v \models \varphi$; therefore, the induction hypothesis gives that $\mathcal{M}^T, \langle \bar{w}'', h_v \rangle \models \varphi$. With this, we have shown that our claim is true and that $\mathcal{M}^T, \langle \bar{w}, h_w \rangle \models \odot_S[\alpha]\varphi$.

(\Leftarrow) We assume that $\mathcal{M}^T, \langle \bar{w}, h_w \rangle \models \odot_S[\alpha]\varphi$. Then for every $L^T \in \text{Choice}_{\alpha}^{\bar{w}}$ such that $\mathcal{M}^T, \langle \bar{w}, h_z \rangle \not\models \varphi$ for some $h_z \in [L^T]_{\alpha}^{\bar{w}'}$ (with $\bar{w} \sim_{\alpha} \bar{w}'$), there exists $L_1^T \in \text{Choice}_{\alpha}^{\bar{w}}$ such that $L^T \prec_s L_1^T$ and if $U = L_1^T$ or $L^T \preceq_s U$, then for every \bar{w}'' such that $\bar{w} \sim_{\alpha} \bar{w}''$, $\mathcal{M}^T, \langle \bar{w}'', h_u \rangle \models \varphi$ for every $h_u \in [U]_{\alpha}^{\bar{w}''}$. Let $N \in \text{Choice}_{\alpha}^{\bar{w}}$ such that $\mathcal{M}, v \not\models \varphi$ for some $v \in \llbracket N \rrbracket_{\alpha}^{\bar{w}'}$ (with $w \approx_{\alpha} w'$). By induction hypothesis, we get that $\mathcal{M}^T, \langle \bar{w}', h_v \rangle \not\models \varphi$. The fact that $w \approx_{\alpha} w'$ implies that $\bar{w} \sim_{\alpha} \bar{w}'$. Since $h_v \in [N^T]_{\alpha}^{\bar{w}'}$, then, we know that there exists $L_1^T \in \text{Choice}_{\alpha}^{\bar{w}}$ such that $N^T \prec_s L_1^T$ and if $U = L_1^T$ or $L_1^T \preceq_s U$, then for every \bar{w}'' such that $\bar{w} \sim_{\alpha} \bar{w}'', \mathcal{M}^T, \langle \bar{w}'', h_u \rangle \models \varphi$ for every $h_u \in [U]_{\alpha}^{\bar{w}''}$. Here, we claim that L_1 is the choice cell that will witness to the fact that $\mathcal{M}, w \models \odot_S[\alpha]\varphi$. In order to show this, we first notice that the second item of **Claim 1** renders that $N \prec_s L_1$. Next, for every w'' such that $w \approx_{\alpha} w''$ (which implies that $\bar{w} \sim_{\alpha} \bar{w}''$) and every $v \in \llbracket L_1 \rrbracket_{\alpha}^{\bar{w}'}, h_v \in [L_1^T]_{\alpha}^{\bar{w}'}$ and therefore $\mathcal{M}^T, \langle \bar{w}'', h_v \rangle \models \varphi$; in this way, we have that for every \bar{w}'' such that $w \approx_{\alpha} w''$ and every $v \in \llbracket L_1 \rrbracket_{\alpha}^{\bar{w}'}$, the induction hypothesis gives that $\mathcal{M}, v \models \varphi$. Finally, let $L \in \text{Choice}_{\alpha}^{\bar{w}}$ such that $L_1 \preceq_s L$. By the second item of **Claim 1**, this last condition implies that $L_1^T \preceq_s L^T$. Take w'

such that $w \approx_\alpha w'$ (which implies that $\bar{w} \sim_\alpha \bar{w}'$), and take $v \in \llbracket L \rrbracket_\alpha^{\bar{w}}$. We know that $h_v \in [L^T]_\alpha^{\bar{w}'}$, so that the fact that $L_1^T \preceq_s L^T$ implies with our assumption that $\mathcal{M}^T, (w'', h_v) \models \varphi$; therefore, the induction hypothesis gives that $\mathcal{M}, v \models \varphi$. With this, we have shown that our claim is true and that $\mathcal{M}, w \models \odot_S[\alpha]\varphi$.

B.1 Canonical Kripke-estit models

We will show that the proof system Λ is complete with respect to the class of Kripke-estit models (and thus use Proposition 4 to show completeness with respect to the class of epistemic utilitarian bi-valued BT-models). The strategy is to build a canonical structure from the syntax.

Definition 7 (Canonical Structure). *The tuple*

$$\mathcal{M} = \langle W^\Lambda, R_\square, \text{Choice}, \{\approx_\alpha\}_{\alpha \in \text{Ags}}, \text{Value}_\odot, \text{Value}_S, \mathcal{V} \rangle$$

is called a canonical structure for Λ iff

- $W^\Lambda = \{w; w \text{ is a } \Lambda\text{-MCS}\}$.
- R_\square is a relation over W^Λ defined by the following rule: for $w, v \in W^\Lambda$, $w R_\square v$ iff for every φ , $\square\varphi \in w \Rightarrow \varphi \in v$. For $w \in W^\Lambda$, the set $\{v \in W^\Lambda; w R_\square v\}$ is denoted by \bar{w} .
- Choice is a function that assigns to each α and \bar{w} a subset of $\mathcal{P}(\bar{w})$, which will be denoted by $\text{Choice}_\alpha^{\bar{w}}$ and defined as follows: let $R_\alpha^{\bar{w}}$ be a relation on \bar{w} such that for $w, v \in W^\Lambda$, $w R_\alpha^{\bar{w}} v$ iff for every φ , $[\alpha]\varphi \in w \Rightarrow \varphi \in v$; if we take $\text{Choice}_\alpha^{\bar{w}}(v) = \{u \in \bar{w}; v R_\alpha^{\bar{w}} u\}$, then we set $\text{Choice}_\alpha^{\bar{w}} = \bigcup_{v \in \bar{w}} \text{Choice}_\alpha^{\bar{w}}(v)$.
- For each $\alpha \in \text{Ags}$, \approx_α is an epistemic relation on W^Λ given by the following rule: for $w, v \in W^\Lambda$, $w \approx_\alpha v$ iff for every φ , $K_\alpha\varphi \in w \Rightarrow \varphi \in v$.
- Value_\odot and Value_S are defined as follows. For $\alpha \in \text{Ags}$ and $w \in W^\Lambda$, we first define two sets:

$$\Sigma_\alpha^w = \{[\alpha]\varphi; \odot[\alpha]\varphi \in w\}$$

$$\Gamma_\alpha^w = \{K_\alpha\varphi; \odot_S[\alpha]\varphi \in w\}.$$

If we take $\Sigma^w = \bigcup_{\alpha \in \text{Ags}} \Sigma_\alpha^w$ and $\Gamma^w = \bigcup_{\alpha \in \text{Ags}} \Gamma_\alpha^w$, then we define the deontic functions by

$$\text{Value}_\odot(w) = \begin{cases} 1 & \text{iff } \Sigma^w \subseteq w \\ 0 & \text{otherwise.} \end{cases}$$

$$\text{Value}_S(w) = \begin{cases} 1 & \text{iff } \Gamma^w \subseteq w \\ 0 & \text{otherwise.} \end{cases}$$

- \mathcal{V} is the canonical valuation, defined such that $w \in \mathcal{V}(p)$ iff $p \in w$.

Proposition 5. *The canonical structure \mathcal{M} for Λ is a Kripke-estit model.*

Proof. We need to show that the tuple $\langle W^\Lambda, R_\square, \text{Choice}, \{\approx_\alpha\}_{\alpha \in \text{Ags}}, \text{Value}_\odot, \text{Value}_S \rangle$ is a Kripke-estit frame, which amounts to showing that the tuple validates the five items in the definition of Kripke-estit models.

1. It is clear that R_\square is an equivalence relation, since Λ includes the **S5** axioms for \square .
2. Since Λ includes the **S5** axioms for $[\alpha]$ (for each $\alpha \in \text{Ags}$), we have that $R_\alpha^{\bar{w}}$ is an equivalence relation for each $\alpha \in \text{Ags}$ and $w \in W^\Lambda$. Moreover, since Λ includes $\square p \rightarrow [\alpha]p$ as an axiom schema, we have that $R_\alpha \subseteq R_\square$. In this way, **Choice** indeed assigns to each α and \bar{w} a partition of \bar{w} . Now, we must verify that \mathcal{M} validates the constraint $(\text{IA})_K$. In order to do that, we need two intermediate results:

- a) For a fixed $w_* \in W^\Lambda$, we have that $w \in \bar{w}_*$ iff $\{\square\psi; \square\psi \in w_*\} \subseteq w$.
 (\Rightarrow) Let $w \in \bar{w}_*$ (which means that $w_* R_\square w$). Take φ a formula of \mathcal{L}_{KO} such that $\square\varphi \in w_*$. Since w_* is closed under *Modus Ponens*, the (4) axiom for \square implies that $\square\square\varphi \in w_*$ as well. Therefore, by the definition of R_\square , we get that $\square\varphi \in w$.
 (\Leftarrow) We assume that $\{\square\psi; \square\psi \in w_*\} \subseteq w$. Take φ a formula of \mathcal{L}_{KO} such that $\square\varphi \in w_*$. By our assumption, we get that $\square\varphi \in w$. Since w is closed under *Modus Ponens*, the (T) axiom for \square implies that $\varphi \in w$ as well. In this way, we have that the fact that $\square\varphi \in w_*$ implies that $\varphi \in w$, which means that $w_* R_\square w$ and $w \in \bar{w}_*$.
- b) For a fixed $w_* \in W^\Lambda$ and $s : \text{Ags} \rightarrow \mathcal{P}(\bar{w}_*)$ a function that maps α to a member of $\text{Choice}_{\alpha}^{\bar{w}_*}$ such that $v_{s(\alpha)} \in s(\alpha)$, we have that $w \in s(\alpha)$ iff $\Delta_{s(\alpha)} = \{[\alpha]\psi; [\alpha]\psi \in v_{s(\alpha)}\} \subseteq w$.
 (\Rightarrow) Let $w \in s(\alpha)$ (which means that $v_{s(\alpha)} R_\alpha^{\bar{w}_*} w$). Take φ a formula of \mathcal{L}_{KO} such that $[\alpha]\varphi \in v_{s(\alpha)}$. Since $v_{s(\alpha)}$ is closed under *Modus Ponens*, the (4) axiom for $[\alpha]$ implies that $[\alpha][\alpha]\varphi \in v_{s(\alpha)}$ as well. Therefore, by definition of $R_\alpha^{\bar{w}_*}$, we get that $[\alpha]\varphi \in w$.
 (\Leftarrow) We assume that $\Delta_{s(\alpha)} = \{[\alpha]\psi; [\alpha]\psi \in v_{s(\alpha)}\} \subseteq w$. Take φ a formula of \mathcal{L}_{KO} such that $[\alpha]\varphi \in v_{s(\alpha)}$. By our assumption, we get that $[\alpha]\varphi \in w$. Since w is closed under *Modus Ponens*, the (T) axiom for $[\alpha]$ implies that $\varphi \in w$ as well. In this way, we have that the fact that $[\alpha]\varphi \in v_{s(\alpha)}$ implies that $\varphi \in w$, which means that $v_{s(\alpha)} R_\alpha^{\bar{w}_*} w$ and $w \in s(\alpha)$.

Next, we will show that for a fixed $w_* \in W^\Lambda$ and $s : \text{Ags} \rightarrow \mathcal{P}(\bar{w}_*)$ a function that maps α to a member of $\text{Choice}_{\alpha}^{\bar{w}_*}$ such as in item b) above, we have that $\bigcup_{\alpha \in \text{Ags}} \Delta_{s(\alpha)} \cup \{\square\psi; \square\psi \in w_*\}$ is Λ -consistent.

First, we will show that $\bigcup_{\alpha \in \text{Ags}} \Delta_{s(\alpha)}$ is consistent. Suppose that this is not the case. Then there exists a set $\{\varphi_1, \dots, \varphi_n\}$ of formulas of \mathcal{L}_{KO} such that $[\alpha_i]\varphi_i \in v_{s(\alpha_i)}$ for every $1 \leq i \leq n$ and

$$\vdash ([\alpha_1]\varphi_1 \wedge \dots \wedge [\alpha_n]\varphi_n) \rightarrow \perp. \quad (1)$$

Without loss of generality, we assume that $\alpha_i \neq \alpha_j$ for all $j \neq i$ such that $j, i \in \{1, \dots, n\}$ —this assumption hinges on the fact that any stit operator distributes over conjunction. Notice that the fact that $[\alpha_i]\varphi_i \in v_{s(\alpha_i)}$ for every $1 \leq i \leq n$ implies that $\Diamond[\alpha_i] \in w_*$ for every $1 \leq i \leq n$. Since w_* is closed under conjunction, we also have that $\Diamond[\alpha_1]\varphi_1 \wedge \dots \wedge \Diamond[\alpha_n]\varphi_n \in w_*$.

By the independence-of-agents axiom, we have that

$$\vdash \Diamond[\alpha_1]\varphi_1 \wedge \dots \wedge \Diamond[\alpha_n]\varphi_n \rightarrow \Diamond([\alpha_1]\varphi_1 \wedge \dots \wedge [\alpha_n]\varphi_n). \quad (2)$$

Therefore, equations (2) and (1), imply that

$$\vdash \Diamond[\alpha_1]\varphi_1 \wedge \cdots \wedge \Diamond[\alpha_n]\varphi_n \rightarrow \Diamond\perp. \quad (3)$$

But this is a contradiction, since we had seen that $\Diamond[\alpha_1]\varphi_1 \wedge \cdots \wedge \Diamond[\alpha_n]\varphi_n \in w_*$, and w_* is a Λ -MCS. Therefore, $\bigcup_{\alpha \in Ags} \Delta_{s(\alpha)}$ is consistent. Next, we show that the union $\bigcup_{\alpha \in Ags} \Delta_{s(\alpha)} \cup \{\Box\psi; \Box\psi \in w_*\}$ is also consistent. Suppose that this is not the case. Since $\bigcup_{\alpha \in Ags} \Delta_{s(\alpha)}$ is consistent, there must exist sets $\{\varphi_1, \dots, \varphi_n\}$ and $\{\theta_1, \dots, \theta_m\}$ of formulas of \mathcal{L}_{KO} such that $[\alpha_i]\varphi_i \in v_{s(\alpha_i)}$ for every $1 \leq i \leq n$, $\Box\theta_i \in w_*$ for every $1 \leq i \leq m$, and

$$\vdash ([\alpha_1]\varphi_1 \wedge \cdots \wedge [\alpha_n]\varphi_n) \wedge (\Box\theta_1 \wedge \cdots \wedge \Box\theta_m) \rightarrow \perp. \quad (4)$$

Let $\theta = \theta_1 \wedge \cdots \wedge \theta_m$. Since \Box distributes over conjunction, we have that $\Box\theta = \Box\theta_1 \wedge \cdots \wedge \Box\theta_m$, where it is important to mention that since w_* is a Λ -MCS, then $\Box\theta \in w_*$. In these terms, (4) implies that

$$\vdash ([\alpha_1]\varphi_1 \wedge \cdots \wedge [\alpha_n]\varphi_n) \rightarrow \neg\Box\theta. \quad (5)$$

Once again, we assume without loss of generality that $\alpha_i \neq \alpha_j$ for all $j \neq i$ such that $j, i \in \{1, \dots, n\}$. Analogous to the procedure we used to show that $\bigcup_{\alpha \in Ags} \Delta_{s(\alpha)}$ is consistent, (5) implies that

$$\vdash \Diamond[\alpha_1]\varphi_1 \wedge \cdots \wedge \Diamond[\alpha_n]\varphi_n \rightarrow \Diamond\neg\Box\theta. \quad (6)$$

This entails that $\Diamond\neg\Box\theta \in w_*$, but this is a contradiction, since the fact that $\Box\theta \in w_*$ implies with the (4) axiom for \Box that $\Box\Box\theta \in w_*$.

Now, let u_* be the Λ -MCS that includes $\bigcup_{\alpha \in Ags} \Delta_{s(\alpha)} \cup \{\Box\psi; \Box\psi \in w_*\}$. By the intermediate result a), it is clear that $u_* \in \overline{w_*}$. By the intermediate result b), it is clear that $u_* \in s(\alpha)$ for every $\alpha \in Ags$. Therefore, we have shown that for a fixed $w_* \in W$, we have that each function $s : Ags \rightarrow \mathcal{P}(\overline{w_*})$ that maps α to a member of $\text{Choice}_{\alpha}^{\overline{w_*}}$ is such that $\bigcap_{\alpha \in Ags} s(\alpha) \neq \emptyset$, which means that \mathcal{M} validates the constraint (IA)_K.

3. Since the axiom system Λ includes the **S5** axioms for K_α for each $\alpha \in Ags$, we have that \approx_α is an equivalence relation for each $\alpha \in Ags$. We must now verify that \mathcal{M} validates the constraints (OAC)_K and (Unif - H)_K.

- i) For (OAC)_K, fix $w_* \in W^\Lambda$ and $\alpha \in Ags$. We assume that $v, u \in \overline{w_*}$ are such that $v \approx_\alpha u$. Let $v' \in \text{Choice}_{\alpha}^{\overline{w_*}}(v)$. This means that $v R_\alpha v'$. We want to show that $v' \approx_\alpha u$, so let φ be a formula of \mathcal{L}_{KO} such that $K_\alpha\varphi \in v'$. By the (4) axiom for K_α , we have that $K_\alpha K_\alpha\varphi \in v'$. Similarly, since all substitutions of the (OAC) axiom lie within v' and it is closed under *Modus Ponens*, we get that $[\alpha]K_\alpha\varphi$ also lies in v' . Since $v' R_\alpha v$, this implies that $K_\alpha\varphi \in v$. Therefore, our assumption that $v \approx_\alpha u$ entails that $\varphi \in u$. With this, we have shown that the fact that $K_\alpha\varphi \in v'$ implies that $\varphi \in u$, which means that $v' \approx_\alpha u$.
- ii) For (Unif - H)_K, fix $w_1, w_2 \in W^\Lambda$. We assume that there exist $v \in \overline{w_1}$ and $u \in \overline{w_2}$ such that $v \approx_\alpha u$. Take $v' \in \overline{w_1}$.

We will show that $u'' = \{\psi; K_\alpha\psi \in v'\} \cup \{\Box\psi; \Box\psi \in u\}$ is consistent. In order to do so, we will first show that $\{\psi; K_\alpha\psi \in v'\}$ is consistent. Suppose for a contradiction that it is not consistent. Then there exists a set $\{\psi_1, \dots, \psi_n\}$ of formulas of \mathcal{L}_{KO} such that $\{\psi_1, \dots, \psi_n\} \subseteq \{\psi; K_\alpha\psi \in v'\}$ and $\vdash \psi_1 \wedge \dots \wedge \psi_n \rightarrow \perp$ (a). Now, the fact that $\{\psi_1, \dots, \psi_n\} \subseteq \{\psi; K_\alpha\psi \in v'\}$ means that $K_\alpha\psi_i \in v'$ for every $1 \leq i \leq n$; by necessitation of K_α and its distributivity over conjunction, we get that (a) implies that $\vdash K_\alpha\psi_1 \wedge \dots \wedge K_\alpha\psi_n \rightarrow K_\alpha\perp$, but this is a contradiction, since v' is a Λ -MCS which, being, includes $K_\alpha\psi_1 \wedge \dots \wedge K_\alpha\psi_n$.

Next, we show that $u'' = \{\psi; K_\alpha\psi \in v'\} \cup \{\Box\psi; \Box\psi \in u\}$ is also consistent. Suppose for a contradiction that it is not consistent. Since $\{\psi; K_\alpha\psi \in v'\}$ is consistent, there must exist sets $\{\varphi_1, \dots, \varphi_n\}$ and $\{\theta_1, \dots, \theta_m\}$ of formulas of \mathcal{L}_{KO} such that $\{\psi_1, \dots, \psi_n\} \subseteq \{\psi; K_\alpha\psi \in v'\}$, $\Box\theta_i \in w_2$ for every $1 \leq i \leq m$, and $\vdash (\varphi_1 \wedge \dots \wedge \varphi_n) \wedge (\Box\theta_1 \wedge \dots \wedge \Box\theta_m) \rightarrow \perp$ (a). Let $\theta = \theta_1 \wedge \dots \wedge \theta_m$ and $\varphi = \varphi_1 \wedge \dots \wedge \varphi_n$. Since \Box distributes over conjunction, we have that $\Box\theta = \Box\theta_1 \wedge \dots \wedge \Box\theta_m$, where it is important to mention that since u is a Λ -MCS, then $\Box\theta \in u$ and $\Box\Box\theta \in u$ as well (\star). In this way, (a) implies that $\vdash \varphi \rightarrow \neg\Box\theta$ and thus that $\vdash \Diamond\varphi \rightarrow \Diamond\neg\Box\theta$ (b). Now, by necessitation of K_α , we get that (b) implies that $\vdash K_\alpha\Diamond\varphi \rightarrow K_\alpha\Diamond\neg\Box\theta$ (c). Notice that the facts that $K_\alpha\varphi_i \in v'$ for every $1 \leq i \leq n$, that K_α distributes over conjunction, and that v' is a Λ -MCS imply that $K_\alpha\varphi = K_\alpha\varphi_1 \wedge \dots \wedge K_\alpha\varphi_n \in v'$. The fact that $v' \in \overline{w_1} = \overline{v}$ implies that $\Diamond K_\alpha\varphi \in v$, so that (c) and the facts that v contains all substitutions of the (*Unif* – *H*) axiom and is closed under *Modus Ponens* entail that $K_\alpha\Diamond\varphi \in v$. Now, this last inclusion implies, with our assumption that $v \approx_\alpha u$, that $\Diamond\varphi \in u$, which by (b) in turn yields that $\Diamond\neg\Box\theta \in u$, contradicting (\star). Therefore, u'' is consistent.

Finally, let u' be the Λ -MCS that includes u'' . It is clear from its construction that $u' \in \overline{u} = \overline{w_2}$ and that $v' \approx_\alpha u'$,

With this, we have shown that \mathcal{M} validates the constraint (*Unif* – *H*) $_K$.

4. Both Value_O and Value_S are well defined functions with range within \mathbb{R} .

As is usual with canonical structures, our objective is to prove the so-called *truth* lemma, which says that for every formula φ of \mathcal{L}_{KO} and every $w \in W^A$, we have that $\mathcal{M}^n, w \Vdash \varphi$ iff $\varphi \in w$. This is done by induction on φ , and the inductive step for each modal operator requires previous results (such as the important *existence* lemmas). In the case of \Box , K_α , and $[\alpha]$, these previous results are standard (Lemma 1 below). However, in the case of $\odot[\alpha]$ and $\odot_S[\alpha]$, the previous results require a lot of work. In Lemmas 2 and 3 below, we pinpoint the main claims that we need to show the inductive step of the truth lemma for the ought operators. Each item in these two lemmas is shown by a series of sub-lemmas, and the reader is advised to carefully go over the full proofs provided in the supplement to the present work.

Lemma 1 (Existence). *Let \mathcal{M} be the canonical Kripke-estit model for Λ . Let $w \in W^A$. For a given formula φ of \mathcal{L}_{KO} , the following hold:*

1. $\Box\varphi \in w$ iff $\varphi \in v$ for every $v \in \overline{w}$.
2. $[\alpha]\varphi \in w$ iff $\varphi \in v$ for every $v \in \overline{w}$ such that $wR_\alpha v$.

3. $K_\alpha \varphi \in w$ iff $\varphi \in v$ for every $v \in W^A$ such that $w \approx_\alpha v$.

Proof. Let $w \in W^A$, and take φ a formula of \mathcal{L}_{KO} .

1. (\Rightarrow) We assume that $\Box \varphi \in w$. If v lies within \bar{w} , this means that $w R_\Box v$ holds. Therefore, $\varphi \in v$.

(\Leftarrow) We work by contraposition. Assume that $\Box \varphi \notin w$. We will show that there is a world in \bar{w} such that φ does not lie within it. For this, let $v' = \{\psi; \Box \psi \in w\}$, which is consistent by virtue of the following argument: suppose for a contradiction that v' is not consistent; then there exists a set $\{\psi_1, \dots, \psi_n\}$ of formulas of \mathcal{L}_{KO} such that $\{\psi_1, \dots, \psi_n\} \subseteq v'$ and $\vdash \psi_1 \wedge \dots \wedge \psi_n \rightarrow \perp$ (a); now, the fact that $\{\psi_1, \dots, \psi_n\} \subseteq v'$ means that $\Box \psi_i \in w$ for every $1 \leq i \leq n$; due to necessitation of \Box and its distributivity over conjunction, we get that (a) implies that $\vdash \Box \psi_1 \wedge \dots \wedge \Box \psi_n \rightarrow \Box \perp$, but this is a contradiction, since w is a Λ -MCS which, being closed under conjunction, includes $\Box \psi_1 \wedge \dots \wedge \Box \psi_n$. Now, we define $v'' = v' \cup \{\neg \varphi\}$, and show that it is also consistent as follows: suppose for a contradiction that it is not consistent; since v' is consistent, we have that there exists a set $\{\psi_1, \dots, \psi_n\}$ of formulas of \mathcal{L}_{KO} such that $\{\psi_1, \dots, \psi_n\} \subseteq v'$ and $\vdash \psi_1 \wedge \dots \wedge \psi_n \wedge \neg \varphi \rightarrow \perp$, which then implies that $\vdash \psi_1 \wedge \dots \wedge \psi_n \rightarrow \varphi$ (b); due to necessitation of \Box and its distributivity over conjunction, we get that (b) implies that $\vdash \Box \psi_1 \wedge \dots \wedge \Box \psi_n \rightarrow \Box \varphi$ (c); but notice that since w is a Λ -MCS, then $\Box \psi_1 \wedge \dots \wedge \Box \psi_n \in w$, so that (c) and the fact that w is closed under *Modus Ponens* entail that $\Box \varphi \in w$, contradicting the initial assumption that $\Box \varphi \notin w$. Finally, let v be the Λ -MCS that includes v'' . On one hand, it is clear from its construction that $\varphi \notin v$ and that $w R_\Box v$. Thus, we have shown that assuming that $\Box \varphi \notin w$ implies the existence of a $v \in \bar{w}$ such that $\varphi \notin v$.

2. (\Rightarrow) We assume that $[\alpha] \varphi \in w$. Let $v \in \bar{w}$ such that $w R_\alpha v$. The definition of R_α straightforwardly gives that $\varphi \in v$.

(\Leftarrow) We work by contraposition. Assume that $[\alpha] \varphi \notin w$. We will show that there is a world v in \bar{w} such that $w R_\alpha v$ and such that φ does not lie within it. For this, let $v' = \{\psi; [\alpha] \psi \in w\}$, which is consistent by a similar argument than the one introduced in the proof of the above item: suppose for a contradiction that v' is not consistent; then there exists a set $\{\psi_1, \dots, \psi_n\}$ of formulas of \mathcal{L}_{KO} such that $\{\psi_1, \dots, \psi_n\} \subseteq v'$ and $\vdash \psi_1 \wedge \dots \wedge \psi_n \rightarrow \perp$ (a); now, the fact that $\{\psi_1, \dots, \psi_n\} \subseteq v'$ means that $[\alpha] \psi_i \in w$ for every $1 \leq i \leq n$; necessitation of $[\alpha]$ and its distributivity over conjunction yield that (a) implies that $\vdash [\alpha] \psi_1 \wedge \dots \wedge [\alpha] \psi_n \rightarrow [\alpha] \perp$, but this is a contradiction, since w is a Λ -MCS which, being closed under conjunction, includes $[\alpha] \psi_1 \wedge \dots \wedge [\alpha] \psi_n$. Now, we define $v'' = v' \cup \{\neg \varphi\}$, and we show that it is also consistent as follows: suppose for a contradiction that it is not consistent; since v' is consistent, we have that there exists a set $\{\psi_1, \dots, \psi_n\}$ of formulas of \mathcal{L}_{KO} such that $\{\psi_1, \dots, \psi_n\} \subseteq v'$ and $\vdash \psi_1 \wedge \dots \wedge \psi_n \wedge \neg \varphi \rightarrow \perp$, which then implies that $\vdash \psi_1 \wedge \dots \wedge \psi_n \rightarrow \varphi$ (b); due to necessitation of $[\alpha]$ and its distributivity over conjunction, we get that (b) implies that $\vdash [\alpha] \psi_1 \wedge \dots \wedge [\alpha] \psi_n \rightarrow [\alpha] \varphi$ (b); but notice that since w is a Λ -MCS, then $[\alpha] \psi_1 \wedge \dots \wedge [\alpha] \psi_n \in w$, so that (b) and the fact that w is closed under *Modus Ponens* entail that $[\alpha] \varphi \in w$, contradicting the initial assumption that $[\alpha] \varphi \notin w$.

Finally, let v be the Λ -MCS that includes v'' . On one hand, it is clear from its construction that $\varphi \notin v$ and that $wR_\alpha v$, by definition of R_α . On the other, axiom (A2) renders that actually $v \in \overline{w}$ (if $\Box\theta \in w$, then $[\alpha]\theta \in w$ and thus $\theta \in v$). Therefore, we have shown that assuming that $[\alpha]\varphi \notin w$ implies the existence of a $v \in \overline{w}$ such that $wR_\alpha v$ and $\varphi \notin v$.

3. (\Rightarrow) We assume that $K_\alpha\varphi \in w$. Let $v \in W^\Lambda$ such that $w \approx_\alpha v$. The definition of \approx_α straightforwardly gives that $\varphi \in v$.

(\Leftarrow) We work by contraposition. Assume that $K_\alpha\varphi \notin w$. We will show that there is a world v in W^Λ such that $w \approx_\alpha v$ and such that φ does not lie within it. For this, let $v' = \{\psi; K_\alpha\psi \in w\}$, which is consistent by an argument similar to the one introduced in the proofs of the above items: suppose for a contradiction that v' is not consistent; then there exists a set $\{\psi_1, \dots, \psi_n\}$ of formulas of \mathcal{L}_{KO} such that $\{\psi_1, \dots, \psi_n\} \subseteq v'$ and $\vdash \psi_1 \wedge \dots \wedge \psi_n \rightarrow \perp$ (a); now, the fact that $\{\psi_1, \dots, \psi_n\} \subseteq v'$ means that $K_\alpha\psi_i \in w$ for every $1 \leq i \leq n$; necessitation of K_α and its distributivity over conjunction yield that (a) implies that $\vdash K_\alpha\psi_1 \wedge \dots \wedge K_\alpha\psi_n \rightarrow K_\alpha\perp$, but this is a contradiction, since w is a Λ -MCS which, being closed under conjunction, includes $K_\alpha\psi_1 \wedge \dots \wedge K_\alpha\psi_n$. Now, we define $v'' = v' \cup \{\neg\varphi\}$ $v'' = v' \cup \{\neg\varphi\}$, and we show that it is also consistent as follows: suppose for a contradiction that it is not consistent; since v' is consistent, we have that there exists a set $\{\psi_1, \dots, \psi_n\}$ of formulas of \mathcal{L}_{KO} such that $\{\psi_1, \dots, \psi_n\} \subseteq v'$ and $\vdash \psi_1 \wedge \dots \wedge \psi_n \wedge \neg\varphi \rightarrow \perp$, which then implies that $\vdash \psi_1 \wedge \dots \wedge \psi_n \rightarrow \varphi$ (b); due to necessitation of K_α and its distributivity over conjunction, we get that (b) implies that $\vdash K_\alpha\psi_1 \wedge \dots \wedge K_\alpha\psi_n \rightarrow K_\alpha\varphi$ (b); but notice that since w is a Λ -MCS, then $K_\alpha\psi_1 \wedge \dots \wedge K_\alpha\psi_n \in w$, so that (b) and the fact that w is closed under *Modus Ponens* entail that $K_\alpha\varphi \in w$, contradicting the initial assumption that $K_\alpha\varphi \notin w$. Finally, let v be the Λ -MCS that includes v'' . On one hand, it is clear from its construction that $\varphi \notin v$ and that $w \approx_\alpha v$, by definition of \approx_α . Thus, we have shown that assuming that $K_\alpha\varphi \notin w$ implies the existence of a $v \in W^\Lambda$ such that $w \approx_\alpha v$ and $\varphi \notin v$.

Lemma 2 (Existence for objective ought). Fix $\alpha \in \text{Ags}$ and $w \in W^\Lambda$.

- (a) For every formula φ of \mathcal{L}_{KO} , $\odot[\alpha]\varphi \in w$ iff $[\alpha]\varphi \in v$ for every $v \in \overline{w} \subseteq W^\Lambda$ such that $\Sigma_\alpha^w \subseteq v$.
- (b) For every $v \in \text{Choice}_\alpha^{\overline{w}}(w)$, $\Sigma_\alpha^w \subseteq w$ iff $\Sigma_\alpha^w \subseteq v$
- (c) $\Sigma_\alpha^w \subseteq w$ iff $\text{Choice}_\alpha^{\overline{w}}(w) \in \text{Optimal}_\alpha^{\overline{w}}$.
- (d) For each $L \in \text{Choice}_\alpha^{\overline{w}} - \text{Optimal}_\alpha^{\overline{w}}$, there exists $L' \in \text{Optimal}_\alpha^{\overline{w}}$ such that $L \prec L'$. Thus, $\mathcal{M}, w \Vdash \odot[\alpha]\varphi$ iff $\forall L \in \text{Optimal}_\alpha^{\overline{w}}, \mathcal{M}, v \Vdash \varphi$ for every $v \in L$.

Proof. (a) **Claim 1.** Let \mathcal{M} be the canonical Kripke-estit model for Λ . We have that the following hold:

- For $w_* \in W^\Lambda$, $w \in \overline{w_*}$ iff $\{\Box\psi; \Box\psi \in w_*\} \subseteq w$.
- For any pair of agents $\alpha, \alpha \in \text{Ags}$ and any formula φ of \mathcal{L}_{KO} , $\odot[\alpha]\varphi \rightarrow \neg \odot[\alpha]\neg\varphi \in w$ for every $w \in W^\Lambda$. We will refer to this property as “consistency of ought-to-do”.

The proof of the first item goes as follows: (\Rightarrow) Let $w \in \overline{w_*}$ (which means that $w_*R_\Box w$). Take φ a formula of \mathcal{L}_{KO} such that $\Box\varphi \in w_*$. Since w_* is closed under

Modus Ponens, the (4) axiom for \Box implies that $\Box\Box\varphi \in w_*$ as well. Therefore, by definition of R_\Box , we get that $\Box\varphi \in w$. (\Leftarrow) We assume that $\{\Box\psi; \Box\psi \in w_*\} \subseteq w$. Take φ a formula of \mathcal{L}_{KO} such that $\Box\varphi \in w_*$. By our assumption, we get that $\Box\varphi \in W$. Since w is closed under *Modus Ponens*, the (T) axiom for \Box implies that $\varphi \in w$ as well. In this way, we have that the fact that $\Box\varphi \in w_*$ implies that $\varphi \in w$, which means that $w_* R_\Box w$ and $w \in \overline{w_*}$.

The proof of the second item is a direct consequence from the fact that, for $\alpha, \beta \in \text{Ags}$ and φ a formula of \mathcal{L}_{KO} , we have that $\vdash \odot[\alpha]\varphi \rightarrow \neg \odot[\beta]\neg\varphi$. This comes from the fact that, for $\alpha, \beta \in \text{Ags}$, $\vdash \odot[\alpha]\varphi \wedge \odot[\beta]\neg\varphi \rightarrow \perp$ —which can be seen by applying axioms (Oic), (IA), and axiom (T) for $[\alpha]$ and $[\beta]$.

Now we proceed with the proof of the main statement.

(\Rightarrow) The definition of Σ_α^w implies straightforwardly that $[\alpha]\varphi \in v$.

(\Leftarrow) We work by contraposition. Suppose that $\odot[\alpha]\varphi \notin w$.

- We first show that Σ_α^w is consistent. Suppose that Σ_α^w is not consistent. Then there is a set $\{\varphi_1, \dots, \varphi_n\}$ of formulas of \mathcal{L}_{KO} such that $\odot[\alpha]\varphi_i \in w$ for every $1 \leq i \leq n$ and $\vdash ([\alpha]\varphi_1 \wedge \dots \wedge [\alpha]\varphi_n) \rightarrow \perp$. This last thing implies that $\vdash ([\alpha]\varphi_1 \wedge \dots \wedge [\alpha]\varphi_{n-1}) \rightarrow \neg[\alpha]\varphi_n$, so that if we take $\psi = \varphi_1 \wedge \dots \wedge \varphi_{n-1}$, then distributivity of $[\alpha]$ over conjunction gives us that $\vdash [\alpha]\psi \rightarrow \neg[\alpha]\varphi_n$. By necessitation of the ought-to-do operator and its axiom (K), we get that

$$\vdash \odot[\alpha]([\alpha]\psi) \rightarrow \odot[\alpha](\neg[\alpha]\varphi_n). \quad (7)$$

First, we notice that since w is a Λ -MCS, then $\odot[\alpha]\psi \in w$ (by distributivity of $\odot[\alpha]$ over conjunction). The fact that $\odot[\alpha]\psi \in w$ implies with axiom (A4) that $\odot[\alpha]([\alpha]\psi) \in w$, which with equation (7) implies that $\odot[\alpha](\neg[\alpha]\varphi_n) \in w$. However, the fact that $\odot[\alpha]\varphi_n \in w$ implies that $\odot[\alpha]([\alpha]\varphi_n) \in w$. Thus, we both have that $\odot[\alpha](\neg[\alpha]\varphi_n) \in w$ and that $\odot[\alpha]([\alpha]\varphi_n) \in w$, which is a contradiction (by consistency of ought-to-do).

- Next, we observe that the union $\Sigma_\alpha^w \cup \{\Box\psi; \Box\psi \in w\}$ is also consistent. Suppose that $\Sigma_\alpha^w \cup \{\Box\psi; \Box\psi \in w\}$ is not consistent. Since Σ_α^w is consistent, we have that there must exist sets $\{\varphi_1, \dots, \varphi_n\}$ and $\{\theta_1, \dots, \theta_m\}$ of formulas of \mathcal{L}_{KO} such that $\odot[\alpha]\varphi_i \in w$ for every $1 \leq i \leq n$, $\Box\theta_i \in w$ for every $1 \leq i \leq m$, and

$$\vdash ([\alpha]\varphi_1 \wedge \dots \wedge [\alpha]\varphi_n) \wedge (\Box\theta_1 \wedge \dots \wedge \Box\theta_m) \rightarrow \perp. \quad (8)$$

Let $\theta = \theta_1 \wedge \dots \wedge \theta_m$. Since \Box distributes over conjunction, we have that $\Box\theta = \Box\theta_1 \wedge \dots \wedge \Box\theta_m$, where it is important to mention that since w is a Λ -MCS, then $\Box\theta \in w$. In these terms, (8) implies that $\vdash ([\alpha]\varphi_1 \wedge \dots \wedge [\alpha]\varphi_n) \rightarrow \neg\Box\theta$. By necessitation of the ought-to-do operator, its axiom (K), and its distributivity over conjunction, we get that

$$\vdash (\odot[\alpha]([\alpha]\varphi_1) \wedge \dots \wedge \odot[\alpha]([\alpha]\varphi_n)) \rightarrow \odot[\alpha](\neg\Box\theta). \quad (9)$$

We have that $\odot[\alpha]([\alpha]\varphi_i) \in w$ for every $1 \leq i \leq n$. Therefore, their conjunction also lies in w , which by (9) entails that $\odot[\alpha](\neg\Box\theta) \in w$. However, we had also observed that $\Box\theta \in w$, which by axiom (4) for \Box implies that $\Box\Box\theta \in w$ and therefore that $\odot[\alpha](\Box\theta) \in w$. We have struck a contradiction, since we cannot have that both $\odot[\alpha](\neg\Box\theta)$ and $\odot[\alpha](\Box\theta)$ lie in w (consistency of ought-to-do). Therefore, $\Sigma_\alpha^w \cup \{\Box\psi; \Box\psi \in w\}$ is consistent.

- Finally, we observe that the union $\Sigma_\alpha^w \cup \{\Box\psi; \Box\psi \in w\} \cup \{\neg[\alpha]\varphi\}$ is also consistent:

First, recall that we are assuming that $\odot[\alpha]\varphi \notin w$. Now, suppose that $\Sigma_\alpha^w \cup \{\Box\psi; \Box\psi \in w\} \cup \{\neg[\alpha]\varphi\}$ is not consistent. Since $\Sigma_\alpha^w \cup \{\Box\psi; \Box\psi \in w\}$ is consistent, then there must exist sets $\{\varphi_1, \dots, \varphi_n\}$ and $\{\theta_1, \dots, \theta_m\}$ of formulas of \mathcal{L}_{KO} such that $\odot[\alpha]\varphi_i \in w$ for every $1 \leq i \leq n$, $\Box\theta_i \in w$ for every $1 \leq i \leq m$, and

$$\vdash ([\alpha]\varphi_1 \wedge \dots \wedge [\alpha]\varphi_n) \wedge (\Box\theta_1 \wedge \dots \wedge \Box\theta_m) \wedge \neg[\alpha]\varphi \rightarrow \perp. \quad (10)$$

Let $\theta = \theta_1 \wedge \dots \wedge \theta_m$. Since \Box distributes over conjunction, we have that $\Box\theta \in w$. Now, (10) implies that

$$\vdash ([\alpha]\varphi_1 \wedge \dots \wedge [\alpha]\varphi_n) \wedge \Box\theta \rightarrow [\alpha]\varphi. \quad (11)$$

By necessitation of the ought-to-do operator, its axiom (K), and its distributivity over conjunction, we get that

$$\vdash (\odot[\alpha]([\alpha]\varphi_1) \wedge \dots \wedge \odot[\alpha]([\alpha]\varphi_n)) \wedge \odot[\alpha](\Box\theta) \rightarrow \odot[\alpha]([\alpha]\varphi). \quad (12)$$

Similarly as above, we have that $\odot[\alpha]([\alpha]\varphi_i) \in w$ for every $1 \leq i \leq n$. Therefore, their conjunction also lies in w . At the same time, $\Box\theta \in w$ implies that $\Box\Box\theta \in w$ (by axiom (4) for \Box), and therefore that $\odot[\alpha](\Box\theta) \in w$. Therefore, (12) entails that $\odot[\alpha]([\alpha]\varphi) \in w$. But axiom (T) for $[\alpha]$, necessitation for the ought-to-do operator, and (K) for the ought-to-do operator then imply that $\odot[\alpha]\varphi \in w$, and this contradicts the previously shown fact that $\odot[\alpha]\varphi \notin w$.

Let u be the Λ -MCS that includes $\Sigma_\alpha^w \cup \{\Box\psi; \Box\psi \in w\} \cup \neg[\alpha]\varphi$. It is clear that $u \in \overline{w}$ (by the first item of **Claim 1**), that $\Sigma_\alpha^w \subseteq u$, and that $[\alpha]\varphi \notin u$.

- (b) (\Rightarrow) Assume that $\Sigma_\alpha^w \subseteq w$, and let $[\alpha]\varphi \in \Sigma_\alpha^w$. By assumption, we have that $[\alpha]\varphi \in w$. Since w is a Λ -MCS, axiom (4) for $[\alpha]$ implies that $[\alpha]([\alpha]\varphi) \in w$ as well. With the fact that $wR_\alpha^w v$, this implies that $[\alpha]\varphi \in v$. Therefore $\Sigma_\alpha^w \subseteq v$ for every $v \in \text{Choice}_\alpha(w)$.
- (\Leftarrow) Analogous.

- (c) **Claim 2.** For $w \in W^\Lambda$, the set $\{u \in \overline{w}; \text{Value}_\odot(u) = 1\}$ is not empty.
(Proof of claim)

Let $w \in W^\Lambda$. We will first show that $\bigcup_{\alpha \in \text{Ags}} \Sigma_\alpha^w$ is consistent. Let $w \in W^\Lambda$. Suppose that $\bigcup_{\alpha \in \text{Ags}} \Sigma_\alpha^w$ is not consistent. Then there must exist a set $\{\varphi_1, \dots, \varphi_n\}$ of formulas of \mathcal{L}_{KO} such that $\odot[\alpha_i]\varphi_i \in w$ for every $1 \leq i \leq n$ and

$$\vdash ([\alpha_1]\varphi_1 \wedge \dots \wedge [\alpha_n]\varphi_n) \rightarrow \perp. \quad (13)$$

Without loss of generality, we assume that $\alpha_i \neq \alpha_j$ for all $j \neq i$ such that $j, i \in \{1, \dots, n\}$ —this assumption hinges on the fact that any stit operator distributes over conjunction. Since w is a Λ -MCS, then the fact that $\odot[\alpha_i]\varphi_i \in w$ for every $1 \leq i \leq n$ implies that $(\odot[\alpha_1]\varphi_1 \wedge \dots \wedge \odot[\alpha_n]\varphi_n) \in w$. Notice then that axiom (Oic) yields that

$$\vdash (\odot[\alpha_1]\varphi_1 \wedge \dots \wedge \odot[\alpha_n]\varphi_n) \rightarrow \Diamond[\alpha_1]\varphi_1 \wedge \dots \wedge \Diamond[\alpha_n]\varphi_n. \quad (14)$$

By axiom (IA), we have that

$$\vdash \Diamond[\alpha_1]\varphi_1 \wedge \cdots \wedge \Diamond[\alpha_n]\varphi_n \rightarrow \Diamond([\alpha_1]\varphi_1 \wedge \cdots \wedge [\alpha_n]\varphi_n). \quad (15)$$

Therefore, equations (14), (15), and (13) imply that

$$\vdash (\Diamond[\alpha_1]\varphi_1 \wedge \cdots \wedge \Diamond[\alpha_n]\varphi_n) \rightarrow \Diamond\perp. \quad (16)$$

But this is a contradiction, since we had that $(\Diamond[\alpha_1]\varphi_1 \wedge \cdots \wedge \Diamond[\alpha_n]\varphi_n) \in w$, and w is a Λ -MCS. Therefore, $\bigcup_{\alpha \in Ags} \Sigma_\alpha^w$ is consistent.

Next, we show that the union $\bigcup_{\alpha \in Ags} \Sigma_\alpha^w \cup \{\Box\psi; \Box\psi \in w\}$ is also consistent. Suppose that this is not the case. Since $\bigcup_{\alpha \in Ags} \Sigma_\alpha^w$ is consistent, there must exist sets $\{\varphi_1, \dots, \varphi_n\}$ and $\{\theta_1, \dots, \theta_m\}$ of formulas of \mathcal{L}_{KO} such that $\Diamond[\alpha_i]\varphi_i \in w$ for every $1 \leq i \leq n$, $\Box\theta_i \in w$ for every $1 \leq i \leq m$, and

$$\vdash ([\alpha_1]\varphi_1 \wedge \cdots \wedge [\alpha_n]\varphi_n) \wedge (\Box\theta_1 \wedge \cdots \wedge \Box\theta_m) \rightarrow \perp. \quad (17)$$

Let $\theta = \theta_1 \wedge \cdots \wedge \theta_m$. Since \Box distributes over conjunction, we have that $\Box\theta = \Box\theta_1 \wedge \cdots \wedge \Box\theta_m$. Since w is a Λ -MCS, then $\Box\theta \in w$. In these terms, (17) implies that

$$\vdash ([\alpha_1]\varphi_1 \wedge \cdots \wedge [\alpha_n]\varphi_n) \rightarrow \neg\Box\theta. \quad (18)$$

Once again, we assume without loss of generality that $\alpha_i \neq \alpha_j$ for all $j \neq i$ such that $j, i \in \{1, \dots, n\}$. Analogous to the procedure we used to show that $\bigcup_{\alpha \in Ags} \Sigma_\alpha^w$ is consistent, (18) implies that

$$\vdash (\Diamond[\alpha_1]\varphi_1 \wedge \cdots \wedge \Diamond[\alpha_n]\varphi_n) \rightarrow \Diamond\neg\Box\theta. \quad (19)$$

This entails that $\Diamond\neg\Box\theta \in w$, but this is a contradiction, since $\Box\theta \in w$ implies with axiom (4) for *square* that $\Box\Box\theta \in w$. Let u be the Λ -MCS that includes $\bigcup_{\alpha \in Ags} \Sigma_\alpha^w \cup \{\Box\psi; \Box\psi \in w\}$. It is clear that $u \in \bar{w}$ and that $\text{Value}_O(u) = 1$. Therefore, the set $\{u \in \bar{w}; \text{Value}_O(u) = 1\}$ is not empty. Therefore, we have shown that **Claim 2.** is true.

We proceed with the proof of the main statement. Recall that $\text{Optimal}_\alpha^{\bar{w}}$ is defined as $\{L \in \text{Choice}_\alpha^{\bar{w}}; \text{there is no } L' \in \text{Choice}_\alpha^{\bar{w}} \text{ such that } L \prec L'\}$.

Let $w \in W^\Lambda$. (\Rightarrow) Let $w \in W^\Lambda$ such that $\Sigma_\alpha^w \subseteq w$. We will show that for every $L \in \text{Choice}_\alpha^{\bar{w}}$ and $S \in \text{State}_\alpha^{\bar{w}}$, $S \cap L \preceq S \cap \text{Choice}_\alpha^{\bar{w}}(w)$, which would straightforwardly imply that $\text{Choice}_\alpha^{\bar{w}}(w) \in \text{Optimal}_\alpha^{\bar{w}}$. For this, it suffices to show that if $\text{Value}_O(u_*) = 1$ for some $u_* \in S \cap L$, then $\text{Value}_O(o) = 1$ for every $o \in S \cap \text{Choice}_\alpha^{\bar{w}}(w)$. Therefore, take $L \in \text{Choice}_\alpha^{\bar{w}}$, $S \in \text{State}_\alpha^{\bar{w}}$, $o \in S \cap \text{Choice}_\alpha^{\bar{w}}(w)$, and $u_* \in S \cap L$ such that $\text{Value}_O(u_*) = 1$. This last thing means that $\bigcup_{\beta \in Ags} \Sigma_\beta^w \subseteq u_*$.

As an intermediate result, we observe that for every $v \in S$, $\bigcup_{\beta \in Ags - \{\alpha\}} \Sigma_\beta^w \subseteq v$: let $v \in S$; since u_*, v both lie in $S \in \text{State}_\alpha^{\bar{w}}$, this means that $v \in \text{Choice}_\beta^{\bar{w}}(u_*)$ for every $\beta \in Ags - \{\alpha\}$, so that item (b) above implies that $\Sigma_\beta^w \subseteq v$ for every $\beta \in Ags - \{\alpha\}$; therefore, $\bigcup_{\beta \in Ags - \{\alpha\}} \Sigma_\beta^w \subseteq v$.

We want to show that $\text{Value}_O(o) = 1$. Since o was taken in $S \cap \text{Choice}_\alpha^{\bar{w}}(w)$, then, on one hand, our intermediate result gives us that $\bigcup_{\beta \in Ags - \{\alpha\}} \Sigma_\beta^w \subseteq o$; on

the other, we have that $o \in \text{Choice}_{\alpha}^{\bar{w}}(w)$ implies by item (b) above that $\Sigma_{\alpha}^w \subseteq o$. With these two facts, we conclude that $\bigcup_{\beta \in \text{Ags}} \Sigma_{\beta}^w \subseteq o$, so that $\text{Value}_{\mathcal{O}}(o) = 1$. (\Leftarrow) We work by contraposition. Suppose that $\Sigma_{\alpha}^w \not\subseteq w$. By item (b) above, this implies that $\Sigma_{\alpha}^w \not\subseteq o$ for every $o \in \text{Choice}_{\alpha}^{\bar{w}}(w)$, so that $\text{Value}_{\mathcal{O}}(o) = 0$ for every $o \in \text{Choice}_{\alpha}^{\bar{w}}(w)$. We will show that there exists an action in $\text{Choice}_{\alpha}^{\bar{w}}(w)$ that dominates $\text{Choice}_{\alpha}^{\bar{w}}(w)$.

From **Claim 2.** we know that the set $\{u \in \bar{w}; \text{Value}_{\mathcal{O}}(u) = 1\}$ is not empty. Take $u_* \in \{u \in \bar{w}; \text{Value}_{\mathcal{O}}(u) = 1\}$. We will show that $\text{Choice}_{\alpha}^{\bar{w}}(u_*)$ is the action that we are looking for, i.e., we will show (\star) that for every state $S \in \text{State}_{\alpha}^{\bar{w}}$, $S \cap \text{Choice}_{\alpha}^{\bar{w}}(w) \leq S \cap \text{Choice}_{\alpha}^{\bar{w}}(u_*)$, and $(\star\star)$ that there exists a state S_* such that $S_* \cap \text{Choice}_{\alpha}^{\bar{w}}(u_*) \not\leq_s S_* \cap \text{Choice}_{\alpha}^{\bar{w}}(w)$. For (\star) , take any $S \in \text{State}_{\alpha}^{\bar{w}}$. We want to show that $S \cap \text{Choice}_{\alpha}^{\bar{w}}(w) \leq S \cap \text{Choice}_{\alpha}^{\bar{w}}(u_*)$. Since for any $o \in S \cap \text{Choice}_{\alpha}^{\bar{w}}(w)$ we have that $\text{Value}_{\mathcal{O}}(o) = 0$, this entails rather straightforwardly that $S \cap \text{Choice}_{\alpha}^{\bar{w}}(w) \leq S \cap \text{Choice}_{\alpha}^{\bar{w}}(u_*)$. For $(\star\star)$, we set $S_{u_*} = \bigcap_{\beta \in \text{Ags} - \{\alpha\}} \text{Choice}_{\beta}^{\bar{w}}(u_*)$. It is clear that $S_{u_*} \in \text{State}_{\alpha}^{\bar{w}}$ and that $u_* \in S_{u_*} \cap \text{Choice}_{\alpha}^{\bar{w}}(u_*)$. Similarly, we have that $(\text{IA})_K$ implies that $S_{u_*} \cap \text{Choice}_{\alpha}^{\bar{w}}(w) \neq \emptyset$. Since for any $o \in S_{u_*} \cap \text{Choice}_{\alpha}^{\bar{w}}(w)$ we have that $\text{Value}_{\mathcal{O}}(o) = 0$, it is clear that $S_{u_*} \cap \text{Choice}_{\alpha}^{\bar{w}}(u_*) \not\leq S_{u_*} \cap \text{Choice}_{\alpha}^{\bar{w}}(w)$. Now, our results (\star) and $(\star\star)$ together render that $\text{Choice}_{\alpha}^{\bar{w}}(w) \prec \text{Choice}_{\alpha}^{\bar{w}}(u_*)$, and this implies that $\text{Choice}_{\alpha}^{\bar{w}}(w) \notin \text{Optimal}_{\alpha}^{\bar{w}}$.

- (d) Let $w \in W^A$ and $L \in \text{Choice}_{\alpha}^{\bar{w}} - \text{Optimal}_{\alpha}^{\bar{w}}$. By **Claim 2.**, there exists $u_* \in \bar{w}$ such that $\text{Value}_{\mathcal{O}}(u_*) = 1$. Firstly, from item (c), we get that $\text{Choice}_{\alpha}^{\bar{w}}(u_*) \in \text{Optimal}_{\alpha}^{\bar{w}}$. The fact that $L \in \text{Choice}_{\alpha}^{\bar{w}} - \text{Optimal}_{\alpha}^{\bar{w}}$ implies with item (b) that $\text{Value}_{\mathcal{O}}(w) = 0$ for every $w \in L$, so that $S \cap L \leq S \cap \text{Choice}_{\alpha}^{\bar{w}}(u_*)$ for every $S \in \text{State}_{\alpha}^{\bar{w}}$ (\star) . On the other hand, if we take S_* to be the unique state in $\text{State}_{\alpha}^{\bar{w}}$ such that $u_* \in S_*$, then it is clear that $S_* \cap \text{Choice}_{\alpha}^{\bar{w}}(u_*) \not\leq S_* \cap L$ $(\star\star)$. Our results (\star) and $(\star\star)$ together render that $L \prec \text{Choice}_{\alpha}^{\bar{w}}(u_*)$.

As for the previous results for \odot_S , we first observe that for $\alpha \in \text{Ags}$, the **S5** axioms for the combination of operators $K_{\alpha}\Box$ are theorems of Λ , which entails that the relation $R_{\Box} \circ \approx_{\alpha}$ is an equivalence relation. For each $w \in W^A$, we will denote by $\alpha_k[w]$ the set $\{v \in W^A; K_{\alpha}\Box\varphi \in w \Rightarrow \varphi \in v\}$. Since the proof of the following lemma is the most important new development of the present work, we include its basic steps (the details are provided in the supplement).

Lemma 3 (Existence for subjective ought). Fix $\alpha \in \text{Ags}$ and $w \in W^A$.

- (a) For every formula φ of \mathcal{L}_{KO} , $\odot_S[\alpha]\varphi \in w$ iff $K_{\alpha}\varphi \in v$ for every $v \in \alpha_k[w]$ such that $\Gamma_{\alpha}^v \subseteq v$.
- (b) For every $v \in \llbracket \text{Choice}_{\alpha}^{\bar{w}}(w) \rrbracket_{\alpha}^{\bar{w}'} (w \approx_{\alpha} w')$, $\Gamma_{\alpha}^w \subseteq w$ iff $\Gamma_{\alpha}^v \subseteq v$.
- (c) $\Gamma_{\alpha}^w \subseteq w$ iff $\text{Choice}_{\alpha}^{\bar{w}}(w) \in \text{S} - \text{Optimal}_{\alpha}^{\bar{w}}$.
- (d) For each $L \in \text{Choice}_{\alpha}^{\bar{w}} - (\text{S} - \text{Optimal}_{\alpha}^{\bar{w}})$, there exists $L' \in \text{S} - \text{Optimal}_{\alpha}^{\bar{w}}$ such that $L \prec_s L'$. Thus, $\mathcal{M}, w \Vdash \odot_S[\alpha]\varphi$ iff $\forall L \in \text{S} - \text{Optimal}_{\alpha}^{\bar{w}}, \forall w'$ such that $w \approx w', \mathcal{M}, v \Vdash \varphi$ for every $v \in \llbracket L \rrbracket_{\alpha}^{\bar{w}'}$.

Proof. (a) (\Rightarrow) We assume that $\odot_S[\alpha]\varphi \in w$. Since w is a Λ -MCS, this implies with axiom (Cl) that $K_{\alpha}\Box \odot_S[\alpha]\varphi \in w$, so that the fact that $v \in \alpha_k[w]$ yields that

$\odot_S[\alpha]\varphi \in v$. The assumption that $\Gamma_\alpha^v \subseteq v$ then entails that $K_\alpha\varphi \in v$. (\Leftarrow) We work by contraposition. Suppose that $\odot_S[\alpha]\varphi \notin w$. This implies that $\odot_S[\alpha]\varphi \notin v$ for every $v \in \alpha_k[w]$, for if there existed $v \in \alpha_k[w]$ such that $\odot_S[\alpha]\varphi \in v$, then $K_\alpha\Box \odot_S[\alpha]\varphi \in v$ by axiom (Cl) and thus $\odot_S[\alpha]\varphi \in w$, contradicting our assumption. Therefore, we take $v_* \in \alpha_k[w]$, where we know that $\odot_S[\alpha]\varphi \notin v_*$. The union $\Gamma_\alpha^{v_*} \cup \{K_\alpha\Box\psi; K_\alpha\Box\psi \in w\} \cup \{\neg K_\alpha\varphi\}$ is consistent. Let u be the Λ -MCS that includes $\Gamma_\alpha^{v_*} \cup \{K_\alpha\Box\psi; K_\alpha\Box\psi \in w\} \cup \{\neg K_\alpha\varphi\}$. We observe that for every $v \in \alpha_k[w]$, $\Gamma_\alpha^w = \Gamma_\alpha^v$. Then we have that $u \in \alpha_k[w]$, that $\Gamma_\alpha^u = \Gamma_\alpha^{v_*} \subseteq u$, and that $K_\alpha\varphi \notin u$.

- (b) Let $w' \in W^\Lambda$ such that $w \approx_\alpha w'$, and let $v \in \llbracket \text{Choice}_\alpha^{\bar{w}}(w) \rrbracket_\alpha^{\bar{w}'}$. (\Rightarrow) Assume that $\Gamma_\alpha^w \subseteq w$. Taking v in $\llbracket \text{Choice}_\alpha^{\bar{w}}(w) \rrbracket_\alpha^{\bar{w}'}$ implies that $w \approx_\alpha v$. Let $K_\alpha\varphi \in \Gamma_\alpha^w$. By assumption, we have that $K_\alpha\varphi \in w$. Axiom (4) for K_α then implies that $K_\alpha K_\alpha\varphi \in w$. Since $w \approx_\alpha v$, $K_\alpha\varphi \in v$ as well. Therefore, we have shown that $\Gamma_\alpha^w \subseteq v$. (\Leftarrow) Analogous.
- (c) (\Rightarrow) Assume that $\Gamma_\alpha^w \subseteq w$. We show that for every $L \in \text{Choice}_\alpha^{\bar{w}}, L \preceq_s \text{Choice}_\alpha^{\bar{w}}(w)$. This comes from the fact that for any w' such that $w \approx_\alpha w'$, if $\text{Value}_S(u_*) = 1$ for some $u_* \in \llbracket L \rrbracket_\alpha^{\bar{w}'} \cap S$, then $\text{Value}_S(o) = 1$ for every $o \in \llbracket \text{Choice}_\alpha^{\bar{w}}(w) \rrbracket_\alpha^{\bar{w}'} \cap S$. (\Leftarrow) We work by contraposition. Assume that $\Gamma_\alpha^w \not\subseteq w$. The set $\{u \in \bar{w}; \text{Value}_S(u) = 1\}$ is not empty. Take $u_* \in \{u \in \bar{w}; \text{Value}_S(u) = 1\}$. It is the case that $\text{Choice}_\alpha^{\bar{w}}(u_*)$ dominates $\text{Choice}_\alpha^{\bar{w}}(w)$ in the order \prec_S .
- (d) Let $L \in \text{Choice}_\alpha^{\bar{w}} - (\text{S} - \text{Optimal}_\alpha^{\bar{w}})$. We know that there exists $u_* \in \bar{w}$ such that $\text{Value}_S(u_*) = 1$. From item (c) above, we get that $\text{Choice}_\alpha^{\bar{w}}(u_*) \in \text{S} - \text{Optimal}_\alpha^{\bar{w}}$. Notice that the fact that $L \in \text{Choice}_\alpha^{\bar{w}} - (\text{S} - \text{Optimal}_\alpha^{\bar{w}})$ implies with items (b) and (c) above that for every w' such that $w \approx_\alpha w'$, $\text{Value}_S(o) = 0$ for every $o \in \llbracket L \rrbracket_\alpha^{\bar{w}'}$. Therefore, it is clear that for every w' such that $w \approx_\alpha w'$ and every $S \in \text{State}_\alpha^{\bar{w}'}$, $\llbracket L \rrbracket_\alpha^{\bar{w}'} \cap S \leq_s \llbracket \text{Choice}_\alpha^{\bar{w}}(u_*) \rrbracket_\alpha^{\bar{w}'} \cap S$ (*). On the other hand, if we take S_* to be the unique state in $\text{State}_\alpha^{\bar{w}}$ such that $u_* \in S_*$, then $\llbracket \text{Choice}_\alpha^{\bar{w}}(u_*) \rrbracket_\alpha^{\bar{w}} \cap S_* \not\leq_s \llbracket L \rrbracket_\alpha^{\bar{w}} \cap S_*$ (**). Our results (*) and (**) together render that $L \prec_s \text{Choice}_\alpha^{\bar{w}}(u_*)$.
with this we conclude the proof.

Lemma 4 (Truth Lemma). *Let \mathcal{M} be the canonical Kripke-estit model for Λ . For every formula φ of \mathcal{L}_{KO} and every $w \in W^\Lambda$, we have that $\mathcal{M}, w \Vdash \varphi$ iff $\varphi \in w$.*

Proof. We proceed by induction on φ . The cases with the boolean connectives are standard. It remains to deal with the modal operators. For formulas involving \Box , $[\alpha]$, and K_α both directions follow straightforwardly from Lemma 1 (1), (2), and (3), respectively. As for the operators for the objective and subjective oughts, we have the following arguments:

- (“ $\odot[\alpha]$ ”)
(\Rightarrow) We work by contraposition. We assume that $\odot[\alpha]\varphi \notin w$. By Lemma 2 (a), we get that there exists $v' \in \bar{w}$ such that $\Sigma_\alpha^w \subseteq v'$ and $[\alpha]\varphi \notin v'$. By Lemma 1 (2), this implies that there exists $v \in \bar{w}$ such that $v' R_\alpha^w v$ and such that $\varphi \notin v$. By induction hypothesis, we get that $\mathcal{M}, v \not\Vdash \varphi$. From Lemma 2 (c), we know that $\text{Choice}_\alpha^{\bar{w}}(v) = \text{Choice}_\alpha^{\bar{w}}(v') \in \text{Optimal}_\alpha^{\bar{w}}$, but this means that there exists

an “action” in $\text{Optimal}_{\alpha}^{\bar{w}}$ —namely $\text{Choice}_{\alpha}^{\bar{w}}(v')$ — such that $v \in \text{Choice}_{\alpha}^{\bar{w}}(v')$ and $\mathcal{M}, v \not\models \varphi$, which implies by Lemma 2 (d) that $\mathcal{M}, w \not\models \odot[\alpha]\varphi$.

(\Leftarrow) We assume that $\odot[\alpha]\varphi \in w$. By Lemma 2 (a), we get that $[\alpha]\varphi \in v'$ for every $v' \in \bar{w}$ such that $\Sigma_{\alpha}^w \subseteq v'$. Take $L \in \text{Optimal}_{\alpha}^{\bar{w}}$. From Lemma 2 (b) we know that $\Sigma_{\alpha}^w \subseteq v$ for every $v \in L$. Therefore, for every $v \in L$, we have that $[\alpha]\varphi \in v$, which implies by axiom (T) for the stit operator that for every $v \in L$, $\varphi \in v$. By induction hypothesis, we get that for every $v \in L$, $\mathcal{M}, v \models \varphi$, so that $\mathcal{M}, w \models \odot[\alpha]\varphi$.

– (“ $\odot_S[\alpha]$ ”)

(\Rightarrow) We work by contraposition. We assume that $\odot_S[\alpha]\varphi \notin w$. By Lemma 3 (a), we get that there exists $v' \in \alpha_k[w]$ such that $\Gamma_{\alpha}^{v'} \subseteq v'$ and $K_{\alpha}\varphi \notin v'$. By Lemma 1 (3), this in turn implies that there exists $v \in W^A$ such that $v' \approx_{\alpha} v$ and such that $\varphi \notin v$. Let $o_v \in \bar{w}$ be such that $v \approx_{\alpha} o_v$ (we know that such o_v exists because condition (US)_K guarantees that the fact that $v' \in \alpha_k[w]$ implies that there exists o_v in \bar{w} such that $v' \approx_{\alpha} o_v$, so that transitivity of \approx_{α} renders that $v \approx_{\alpha} o_v$ as well). We observe that the fact that $o_v \approx_{\alpha} v$ implies that $v \in \llbracket \text{Choice}_{\alpha}^{\bar{w}}(o_v) \rrbracket_{\alpha}^{\bar{w}}$. Now, by induction hypothesis, we get that $\mathcal{M}, v \not\models \varphi$. From Lemma 3 (b) we get that the facts that $\Gamma_{\alpha}^{v'} \subseteq v'$ and $v' \approx_{\alpha} o_v$ imply that $\Gamma_{\alpha}^{o_v} \subseteq o_v$, so that Lemma 3 (c) then implies $\text{Choice}_{\alpha}^{\bar{w}}(o_v) \in S - \text{Optimal}_{\alpha}^{\bar{w}}$. This means that there exists an “action” in $S - \text{Optimal}_{\alpha}^{\bar{w}}$ —namely $\text{Choice}_{\alpha}^{\bar{w}}(o_v)$ — such that $v \in \llbracket \text{Choice}_{\alpha}^{\bar{w}}(o_v) \rrbracket_{\alpha}^{\bar{w}}$ and $\mathcal{M}, v \not\models \varphi$, which implies by Lemma 3 (d) that $\mathcal{M}, w \not\models \odot_S[\alpha]\varphi$.

(\Leftarrow) We assume that $\odot_S[\alpha]\varphi \in w$. By Lemma 3 (a), we get that $K_{\alpha}\varphi \in v$ for every $v \in \alpha_k[w]$ such that $\Gamma_{\alpha}^v \subseteq v$. Take $L \in S - \text{Optimal}_{\alpha}^{\bar{w}}$. From Lemma 3 (b) we know that for every w' such that $w \approx_{\beta} w'$, $\Gamma_{\alpha}^{v'} \subseteq v'$ for every $v' \in \llbracket L \rrbracket_{\alpha}^{\bar{w}'}$. Therefore, for every $v' \in \llbracket L \rrbracket_{\beta}^{\bar{w}'}$, we have that $K_{\alpha}\varphi \in v'$, which implies—by axiom (T) for the knowledge operator—that for every $v' \in \llbracket L \rrbracket_{\alpha}^{\bar{w}'}$, $\varphi \in v'$. By induction hypothesis, we get that for every w' such that $w \approx_{\beta} w'$ and $v' \in \llbracket L \rrbracket_{\alpha}^{\bar{w}'}$, $\mathcal{M}, v' \models \varphi$. Since L was taken as an arbitrary “action” in $S - \text{Optimal}_{\alpha}^{\bar{w}}$, this means that $\mathcal{M}, w \models \odot_S[\alpha]\varphi$, according to Lemma 3 (d).

With this we conclude the proof.

Theorem 1 (Completeness w.r.t. Kripke-estit models). *The proof system Λ is complete with respect to the class of Kripke-estit models.*

Proof. Let φ be a Λ -consistent formula of \mathcal{L}_{KO} . Let Φ be the Λ -MCS including φ . We have seen that the canonical Kripke-estit model is such that $\mathcal{M}, \Phi \models \varphi$.

Theorem 2 (Completeness w.r.t. bi-valued BT-models). *The proof system Λ is complete with respect to the class of epistemic utilitarian bi-valued BT-models.*

Proof. Let φ be a Λ -consistent formula of \mathcal{L}_{KO} . By Theorem 1 we know that there exists a Kripke-estit model \mathcal{M} and a world w in its domain such that $\mathcal{M}, w \models \varphi$. Proposition 4 then ensures that the epistemic utilitarian bi-valued BT-model \mathcal{M}^T associated to \mathcal{M} is such that $\mathcal{M}^T, \langle \bar{w}, h_w \rangle \models \varphi$.